Seattle
Police Department

December 8, 2023

M E M O R A N D U M

SUBJECT: December 15th Report to Court

Under the Court's September 7, 2023, Order Granting in Part and Denying in Part the Parties' Joint Motion to Approve Proposed Agreement on Sustained Compliance (Order), the Seattle Police Department is obligated to report to the Court, by December 15, 2023, its work across three topic areas: (1) status of its implementation of RCW 10.114.011 and RCW 43.102.020, regarding the use of deadly force; (2) efforts to improve data transparency, usability and accessibility; and (3) its plan for identifying and mitigating racial disparities in use of force, crisis intervention, and stops and detentions. In satisfaction of these requirements, SPD reports as follows.

I. Status on Implementation of RW 10.114.011 and RCW 43.102.020.

Pursuant to the terms of the Consent Decree and court-approved manuals and policies developed thereunder, all use of force by Seattle Police officers is investigated and reviewed internally to a degree that varies depending on the level of force used. This appropriately directs resources proportionately based on the severity and risk of any use of force. For serious use of force, up to and including the use of deadly force, the actions of involved officers are investigated by the Force Investigation Team (FIT) – a detective team with "special training on gathering evidence, interviewing witnesses, and exploring avenues of inquiry not merely relating to the moment that the force was applied but also on the events, decisions and tactics that led up to the use of force incident." Dkt. 231, 17. Since the Monitor's First Systemic Assessment, filed in 2015, the Monitor has consistently praised the quality of FIT investigations – from post-incident response through its investigative process and fact-finding – to be sound:

> FIT's investigations are covering all relevant investigative lines of inquiry, probing important issues and attempting to resolve inconsistencies among statements and evidence. In multiple instances, Monitoring Team reviewers saw the quality of SPD's force response and investigation improve immediately upon FIT's arriving at the scene or beginning to investigate the incident. FIT is providing SPD chain of command with fair, thorough, complete, and objective factual records from which to make determinations about whether officer performance involving force is consistent with SPD policy.

The former Monitor determined that "[s]o long as all of FIT's policies and procedures are codified in the FIT Manual … the Monitor believes that FIT will be able to maintain compliance with the requirements of paragraph 118 of the Consent Decree[.]" The current monitor affirmed, in his 2022 Comprehensive Assessment finding sustained compliance, FIT's continued high performance.
While the Consent Decree was explicit in directing internal investigations and review of all use of force, several changes in Washington state law will require a shift in protocol once SPD is relieved of federal oversight in this area. By way of short history:

- In 2017, Washington voters approved Initiative 940 (I-940), which, in addition to imposing additional training requirements and creating an affirmative duty to render aid, called for

independent investigations into any use of deadly force resulting in death, great bodily harm, or substantial bodily harm.

- In 2019, the Washington legislature passed Substitute House Bill 1064, which codified into Chapter 10.114 RCW the provisions of I-940:

  Except as required by a federal consent decree, federal settlement agreement, or federal court order, where the use of deadly force by a peace officer results in death, substantial bodily harm, or great bodily harm, an independent investigation must be completed to inform any determination of whether the use of deadly force met the good faith standard established in RCW 9A.16.040 and satisfied other applicable laws and policies. The criminal justice training commission must adopt rules establishing criteria to determine what qualifies as an independent investigation pursuant to this section.

- In October 2019, the CJTC filed the first of its proposed rules,[3] contemplating that agencies within geographic proximity would enter into agreements to form independent investigation teams (IITs), establishing the membership and training of such teams, and setting forth standards for investigations. As an agency under a consent decree, and thus excepted from the requirements of RCW 10.114, SPD is not currently a member of any IIT.

- In 2021, the legislature passed ESSB 1267, amending RCW 10.114.011 and adding a new chapter to Title 43 RCW (RCW 43.102) establishing in the office of the governor the Office of Independent Investigations (OII) with authority to take on, at its discretion, independent investigations prescribed in RCW 10.114.011 that are presently being conducted by IITs in accordance with WAC 139-12.

At present time, SPD is proceeding on two tracks to ensure compliance with state law once the consent decree is terminated.

(1) SPD has been in discussion with King County's Independent Force Investigation Team (IFIT) about joining that IIT until the OII is up and running and prepared to take new cases. We do not know, and will likely not know before this filing, whether KC IFIT will enter into an agreement with SPD, or if so in what capacity SPD will contribute.

(2) SPD has also been in frequent communication with OII Director Rogoff and his team regarding the eventual transition of this body of work to that office.

The investigations under RCW 10.114 (IFIT) and RCW 43.102 (OII), while broader in scope than a criminal homicide investigation, do not subsume SPD's administrative FIT investigations. However, these laws may have a significant impact on SPD's control and access to evidence. The Consent Decree focused resources on a timely and robust administrative investigation process to ensure fidelity to policy and training and drive departmental learning and evolution based on critical events. SPD has not yet had the opportunity to fully deconflict its current court-approved policies and FIT procedures with the eventual protocols of the independent investigations required, either

under RCW 10.114.011 or Chapter 43.102 (the latter of which, SPD understands, are still in development). Based upon the requirements of state law and the Washington Administrative Code, however, the following are specific areas of SPD policy that will be or, depending on how eventual protocols are written, may be impacted in all cases falling within the purview of state law:

- While SPD may hold a scene until an IIT/OII team arrives, neither FIT nor SPD CSI will have access to the scene until cleared by the independent team. To SPD's current understanding, neither the Office of Police Accountability nor the Inspector General will be permitted to be present. If true, this may bear on OPA investigation timelines and will make SPD reliant on the timing, quality, and thoroughness of the independent scene investigation.

- SPD FIT may no longer have immediate access to the involved or witness officers; SPD's ability to photograph, document equipment, and complete in-person interviews of involved officers before they end their shift, accordingly, may be significantly impacted (FIT Manual). Similarly, SPD may no longer have timely access to third party witnesses.

- Because SPD's access to interviews and evidence will likely be limited while the independent investigations are pending, deadlines for FIT to complete its administrative investigations of force (and for Force Review Board (FRB) consideration of the matter) will need to be adjusted commensurately.

- The Force Review Board (FRB) receives a comprehensive presentation from SPD FIT, as well as the complete evidentiary SPD FIT file, prior to deliberating on a Type III use of force. To the degree the SPD FIT investigation is impacted, the FRB review will be similarly affected.

- SPD policy concerning the release of body worn camera footage and other objective evidence within 72 hours of the incident may be impacted. SPD fully intends to release any such evidence it has access to, and if access is not permitted, will release a statement to that effect.

SPD appreciates the early collaboration with the OII to ensure that, to the extent feasible within that office's statutory remit, SPD is able to continue with its administrative investigation and review process and commitments to public transparency.

SPD FIT will continue to deploy consistent with current policy to all Type III use of force cases that do not fall within the independent investigation criteria.

II. Data Transparency, Usability, and Accessibility

The evolution of SPD's data systems, from the fractured, patched-together siloes noted by the Department of Justice in its 2011 Findings Letter to what the Monitor described, in his 2022 Stops and Detention Assessment, as a "data infrastructure and analytical capacity [that allows SPD] to conduct rigorous analysis on an ongoing basis to support evidence-based management practices," is well documented in the case record.[6] As detailed throughout the record, SPD has worked hard to improve its data collection, use its data as a matter of course, and to be transparent in its work and outcomes. As a result of dedicated system improvements over the past decade, current data systems are mature, data collection is without gaps, and the breadth of data is used routinely, internally and for public presentation. Indeed, as the Monitor noted in his May 2022 Comprehensive Assessment, SPD's transformation from an agency "lacking meaningful internal data in many respects at the beginning of this process to now producing extensive public data and dashboards on areas of public interest has been a notable, importance achievement over the course of the Consent Decree[.]"

Data and a commitment to transparency are central to the open data strategy at SPD and tools are designed to build a shared understanding of how police service is administered by SPD. In addition to context, methodology, and analysis available from the Information and Data pages,[7] the open data portal contains comprehensive metadata pages describing where the data come from, what they represent, how they are updated, what each field means and how it is best/appropriately used. Cautions and caveats are included. These pages additionally host infographic short videos describing how the business of police materializes in the data. Currently, Calls for Service, Police Reports and Use of Force videos are already complete; a Missing Persons video in progress and videos detailing interactions with persons in crisis, *Terry* stops, arrests, and complaints will follow in the near future.

Since the launch of the first version of the Data Analytics Platform in 2017, SPD has continually generated insights into process, data, and infrastructure continuity and quality. The Data Governance program actively documents, analyzes, and tracks business processes, infrastructure, and engineering improvements to render these data accessible for ad hoc and dashboard reporting. To date, more than 600 Data Governance Activity Log (DGAL) items have been logged and are integrated with other data engineering and analytical work tracked through the Department's agile project administration platform, Jira. Still, we invite our partners, stakeholders, and community to ask questions, point out errors and generally join the conversation to continuously improve the delivery of police service in Seattle.

Over the past year, SPD has taken several additional significant steps to take its analytic capacity, usability, and transparency to the next level.

- SPD implemented DAP Version 2.0, moving it from on-premises servers to a cloud environment, which significantly increases SPD's and the City's capacity to provide evidence-based public safety analysis and technology solutions by (1) streamlining access to technology, including applications and computing services; (2) providing easier methods of collaborating with the public, as well as public and private academic institutions; and (3) allows for a hypothesis-driven approach to analytics. The cloud environment allows for easier ingestion of data, comes with an array of native Amazon Web Service tools that can be readily deployed in that environment, and allows the department to choose from a range of instance types to meet processing/modeling needs.

- In July 2023, SPD consolidated its data analytics teams into a single bureau, merging what had been two siloed teams operating under bifurcated leadership (Performance Analytics and Research, which focuses on internally facing data relating to officer and department activities, and Data Driven, which focuses on crime data analysis and reporting) into one consolidated team. Operating under a commitment to responsible applications of data and analytics consistent with the White House's recently released guidance on advancing equitable data[8] and leveraging this combined expertise, including advanced degrees up to and including the doctoral level, SPD has completely reformatted SeaStat (SPD's model of CompStat) to integrate advanced analytics, such as change-point detection, to support an evidence-based approach to testing and tracking methods of targeted public safety interventions.

- While SPD continues to strive for a perfect score, SPD is proud to have recently achieved the highest score of any agency in the nation on the Vera Institute of Justice Police Data Transparency Index.

To further the usability and accessibility of SPD's data, SPD sought, and was recently awarded, a grant from the Bureau of Justice Assistance's National Training and Technical Assistance Center (NTTAC) to update and upgrade SPD's public-facing dashboards. SPD appreciates the partnership of the Office of Inspector General,

which recently completed a comprehensive assessment of SPD's dashboards, for providing a solid roadmap for application of this grant.

III. Identifying and Mitigating Racial Disparities in Use of Force, Crisis Intervention, and Stops and Detentions

Assessing disparities in policing data is, as the Monitoring Team has consistently acknowledged, a complex exercise. A standard practice that many may default to, simply comparing demographics represented in police activity across census-based population demographic data, is of limited value if the goal is to guide management practices:

> One common disparity analysis involves comparing police data on topics like stops or uses of force against population statistics to examine whether police actions are impacting certain demographics in a disproportionate fashion. Population-based analyses present insights but also do not, by themselves, tell a complete story regarding disparity or potential bias, since other sociological factors may impact policing disparities as they do in other areas of society. Consequently, population-based comparisons do "not tell us much about what is driving disparity," as noted by the previous Monitoring Team. For example, this assessment will show that SPD use of force and stops disproportionately impact certain minority groups in Seattle, but these population-based conclusions cannot identify to what degree these disparities result specifically from SPD apart from broader sociological forces. Certainly, the limitations of population-based disparities do not mean that such disparities lack meaning. Rather, they are an important way of reviewing police activity, but it is likewise important to remain cognizant of their limitations in factoring in other potentially relevant social forces or explaining why specifically disparities are occurring – or what specifically can be done to address the identified disparities.

The Monitor also cited a study from the Center for Policing Equity on this point:

> For example, the Center for Policing Equity (CPE) published a 2021 report comparing SPD's use of force and stop practices against Seattle's population. CPE is national leader in assessing and addressing these very issues. Like previous Monitoring Team assessments, CPE's report found disparities in stop trends that once again prompted community concerns regarding the racial impacts of policing in Seattle. CPE contextualized what these findings meant up front in their report:
>
> > While findings of racial disparities are always reason for concern, they are not necessarily attributable to decisions or practices by law enforcement. In other words, observed racial disparities do not necessarily indicate that officers have prejudiced beliefs or that they have even engaged in discriminatory behavior. Crime, poverty, institutional neglect, and a host of other factors may drive law enforcement's disparate contacts with and other behaviors toward various racial groups. These factors do not mean disparities are not a concern, just that those seeking to address the concern must focus on all of the factors that produce them—including, but not limited to, the policies and behaviors of law enforcement.

This important framing was also specifically recognized by the Court in its Order directing this response:

> The court recognizes that it is important to acknowledge and understand the extent to which disparities in policing arise from disparities upstream of police interactions. These upstream disparities, however, call for whole-of-government intervention that is beyond the scope of the City's obligations under the Consent Decree. Order, 10, fn. 4.

SPD employs two of the most sophisticated analytical methods for the identification and mitigation of collateral harms, including racially disparate impact. First, as SPD described in a 2019 report to the court, and as is described in the Monitor's 2021 Semi-Annual Report, SPD began testing the use of propensity score matching – a quasi-experimental technique that approximates experimental balance between groups such that one variable (here, subject race) can be isolated – to observe disparity in certain discretionary actions (such as the decision to frisk a subject during a *Terry* stop) over time. This observability allows the department to better understand any trends at a high level of resolution (precinct-sector-month), such that the department (and supervisors/commanders) can focus on identifying and mitigating any operational practices that might be leading to otherwise unexplainable disparity in outcomes of police contact. Continuous improvement of this evolved propensity score weighting approach is being developed to make optimal use of the available data at increasingly more detailed levels of analysis.

Second, using automatic vehicle location "pings" from patrol vehicles, SPD began testing a spatial analysis approach to creating a measure of the ratio between community need for police in a specific area and the corresponding level of police presence in order to identify areas of over- and under-policing relative to service demand. This not only allows supervisors greater visibility into where officers are spending discretionary time – which can be both an officer safety and community trust issue – but allows the department to better monitor efficiency in patrol activity. (A more comprehensive, peer-reviewed assessment of SPD's approach is attached as Attachment A.)

As the Monitor noted in his 2022 Comprehensive Assessment, "[f]ew, if any, law enforcement agencies in the United States have built or maintain the internal capacity to produce ongoing disparity analyses at [SPD's] level of rigor and sophistication." In its last reporting on these efforts, SPD was at a "proof of concept" phase with both methodologies and, consistent with its commitment to ensuring rigor in its work, had partnered with RTI International's Center for Policing Research and Investigative Science to validate these approaches. A copy of that evaluation, issued March 2023 and confirming the viability of both approaches, is attached as Attachment B. Leveraging the expanded computing capacity made possible by transitioning the DAP to a cloud environment, SPD will be extending these approaches to monitoring and assessing disparity across additional sources.

Highlighting SPD's work in this area, the Monitor's 2022 Comprehensive Assessment expressed hope that SPD's work would "catalyze the City of Seattle to identify systematically the types of activity that lead to disproportionate impacts and explore potential alternative responses that might reduce or eliminate such disparities." While SPD has continued to build upon those capacities that are within SPD's operational and analytic control, it is also important to call out the important work that has been done at the City level to build out non-law enforcement capacity to address low-acuity calls often driven by those factors – poverty, substance use disorder, mental health crises – upstream of police contacts that can in turn drive disparity in policing data. SPD looks forward to partnering with this new department – Community Assisted Response and Engagement – to identify yet further opportunities to mitigate harms that can be perpetuated throughout criminal justice and social service systems. This is an important step in a "whole-of-government intervention" directed at providing services upstream of police intervention and SPD is fully supportive of these efforts, which should lead to better outcomes at the intersection of public health and safety.

ATTACHMENT 1

March 2023

Seattle EAQ Evaluation Final Report

Prepared for

Seattle Police Department
Loren T. Atherley
Seattle Police Department
P.O. Box 34986
Seattle, WA 98124-4986
Loren.atherley@seattle.gov

Prepared by

Sean E. Wire
Elise P. Kratzer
Heather Prince, PhD
RTI International
3040 E. Cornwallis Road
P.O. Box 12194
Research Triangle Park, NC 27709

RTI Project Number 0218245.000.001

RTI INTERNATIONAL

Contents

Executive Summary

Seattle's Equity, Accountability, and Quality (EAQ) initiative is a holistic risk-management approach that aims to actively manage the balance between crime control and civil liberties and examine the total cost of ownership of public safety. The EAQ model replicates the framework of a police performance management system (CompStat)-style management meeting, using novel metrics developed by the Seattle Police Department to demonstrate compliance with the consent decree and in an attempt to reflect organizational health based on equity, accountability, and quality measures. This report draws the following conclusions about each component in the EAQ.

1. The **post-stop equity component** uses a propensity score-matching approach as an elegant way to isolate racial bias where causal experimentation is not possible. This metric leverages proven methodologies and existing data collection efforts to provide a reasonable measure of racial disparity in officer decision making. We recommend that this component proceed as an integral part of the equity component of EAQ, with additional concurrent efforts to improve data quality and the subsequent accuracy of this measure.

2. The **location-based resource accountability component** satisfies its intended goal of creating a dynamic measure of resident need and police service. The data sources for both resident need (calls for service) and officer location (Automated Vehicular Locator) are both appropriate and available, and the metric itself relies on well-established spatial statistics to identify a novel need. Due to the discussed limitations to geographic specificity, and the reliance on officer location over behavior, we do not view this metric as a definitive measure of over-policing. However, it does serve as a valid start to conversations about why observed disparities occur and how to iteratively improve service equity from there. We recommend this component proceed as a valuable part of the EAQ forum.

3. Due to the recent shift in the interaction quality metric from body-video assessment to **post-interaction surveys**, there are unknowns related to implementation. The use of survey techniques to evaluate perceptions of police performance is a well-established methodology, here facilitated by improved distribution and survey targeting. We believe that, in its current form, the metric's use of Net Promoter Scores provides an effective benchmark for police performance as a whole, with some current technical limitations to evaluating subjective

interaction quality. We can offer only initial recommendations, although its value in the EAQ may improve as future efforts expand survey content and generalizability.

1. Introduction

1.1 Background

In 2012, the City of Seattle entered into a settlement agreement with the U.S. Department of Justice (DOJ) as the result of a DOJ investigation into the Seattle Police Department's (SPD's) use of force practices and concerns about biased policing. Although the investigation did not find that SPD officers engaged in biased policing, it noted concerns regarding racial disparities in outcomes. With these findings, SPD began working with the DOJ Office of the Inspector General to establish strategies and metrics to demonstrate progress toward compliance with post-Federal Consent Decree operations. Through this strategy, SPD is working to extend and sustain progress toward a more equitable delivery of police service established under its consent decree by regularly collecting and reporting Equity, Accountability, and Quality (EAQ) measures that support a focus on continuous improvement. SPD has upgraded its data warehouse and processing infrastructure to provide near real-time patterns of disparity and evolve its general understanding of the collateral harms associated with the delivery of police service. With this, SPD is establishing a CompStat-style forum that will focus on the continuous monitoring of three established EAQ measures and will look at operationalized measures for disparate outcomes, under- and over-policing of communities, and service quality for awareness, mitigation, and continuous improvement. Overall, this process continually maintains focus and progress on the consent decree reforms. SPD has partnered with RTI International to serve as its quality assurance and evaluation partner for implementing the EAQ process.

RTI is a private, nonprofit research organization with the capabilities, infrastructure, and review systems to manage and complete complex projects. RTI's Center for Policing Research and Investigative Science actively partners with law enforcement agencies across the country, with an emphasis on providing rigorous, data-driven results that have direct implications for the field. In September 2021, RTI began documenting the planned EAQ methodologies and technology through a series of meetings with the City of Seattle, SPD, and relevant partners to understand the development and implementation of methodological approaches. For each method, RTI staff reviewed relevant research and literature relating to the foundational hypotheses for the methods and technology presented and consulted internal and external experts on best practices and feasibility of the methodology.

1.2 Format

Broadly, this report serves as a direct follow-up to the pre-implementation report submitted to the SPD in December 2021. During this intervening time, SPD has refined the methodological components of the EAQ, taking into account initial concerns and considerations raised by RTI during the first evaluation. This report will revisit the methodologies of the following three EAQ components, providing an updated evaluation of the proposed application and methodology:

- **Terry Frisk Equity Evaluation**: To measure the level of racial equity in officers' decisions to frisk during a Terry stop, SPD is using Propensity Score Matching (PSM) to determine whether observed disparity in post-stop outcomes can be attributed to the perceived race of the subject, with the expectation that this approach can translate to other areas of service equity.
- **Location-Based Resource Accountability**: To look at areas of over- and under-policing relative to demand, SPD is using spatial analysis to map areas of community need and Automated Vehicular Locator (AVL) data to measure police presence.
- **Officer Service Quality Assessment**: To assess the quality of interactions between officers and community members more frequently, SPD is utilizing SPIDR Tech to obtain feedback from community members following interactions and contribute to a composite rating of interaction quality for the department.

For this report to serve as a comprehensive evaluation, there is some intentional overlap with the pre-implementation report, wherein the project description, logical evaluation and methodological validation are included and updated where necessary. SPD has made concerted efforts to consider RTI's initial criticisms and opportunities for improvement of the proposed methodologies, where possible. We will highlight the resolution of relevant considerations for each component in this report. Where no such clarification or change has been made, we provide a list of outstanding considerations that may be relevant to how the metrics are employed or may be improved in the future. Finally, for each of the three components, we provide a summary judgment of whether and how SPD should proceed in integrating the component into the overall EAQ framework.

1.3 Scope

In this report, we intend to raise all relevant considerations as justified by the existing literature and expert input. However, there are two key restrictions to the scope of this report. First, this report is limited to a logical and methodological assessment of the three components. Regular EAQ meetings were not yet active during the evaluation period, so we are unable to advise on operational parameters, such as the appropriate cadence of the meetings, how to best leverage the metrics, or how these new evaluative criteria might be communicated within and outside the agency.

Second, our assessment of each component of the EAQ is restricted to its intended use as a system-level metric for police performance. Throughout this evaluative process, our inquiries have been focused on evaluating the EAQ outcomes at this high level of aggregation. In practice, however, each of the three metrics has the potential to be used for more detailed explorations of outcomes at the officer level. There are additional methodological and operational considerations about the appropriateness of applying these methodologies to a different unit of analysis. Therefore, we recommend further consideration before proceeding with these metrics at the officer level.

2. Component 1: Terry Frisk Equity Evaluation

2.1 Evaluating the Premise

The primary goal of this EAQ component is to create a dynamic and high-resolution measure of the level of racial equity in officers' decisions to frisk during a Terry stop to determine whether differences in post-stop outcomes can be attributed to the race of the subject. Terry stops were selected as a convenient and accessible sampling point, but the intention is to measure officers' differential perception of dangerousness that can characterize service equity across a range of scenarios. They will use a PSM approach, whereby situational and demographic variables that SPD recorded will serve as controls, so that any remaining difference can, in theory, be attributed to the race of the subject. The sourcing of adequate data and the ability to process those data efficiently will determine the appropriateness of this proposed methodology.

The proposed data source for this component is the universe of recorded Terry-stop contacts to inform the model, supplemented with computer-aided dispatch data. This incident-level metadata provides measures on officer and subject demographics, situational dimensions surrounding each stop, and abstractions of what a reasonable officer might know prior to their decision to frisk. Barring any non-systematic missing data issues, this internal data source is the only viable way to capture the predicates and outcomes of each police stop.

Because true experimentation is not possible in this case, PSM presents a robust quasi-experimental alternative that is logically consistent with the goals of this EAQ metric. As explored in depth in the pre-implementation report, we find that both the availability of the data and the approach to analysis satisfy the research premise, although we caution here against treating this metric as a comprehensive measure of racial disparity.

2.2 Validating the Methodology

This component uses propensity scores to approximate equivalence between groups and isolate race as responsible for any observed disparity. PSM, which is a quasi-experimental method that matches individuals on demographic and situational similarity, is capable of causal inference in theory (Rosenbaum & Rubin, 1983; Dehejia & Wahba, 2002). In reality, however, the inability to account for unknown unknowns prevents balancing on key differences beyond the treatment effect

(Govindasamy, 2016; King & Nielsen, 2019). The approach SPD proposed assigns propensity scores as weights to account for selection assignment differences (Olmos & Govindasamy, 2015).

New York City has successfully used this approach to assess racial disparity in post-stop outcomes (Levchak, 2021). This study echoes previous research that finds these methods reasonably approximate a randomized experimental design to allow for estimating causal effects. The current approach is also modeled on previous efforts in Seattle. In the *Disparity Review: Part 1* (Seattle Police Department, 2019), a similar methodology was used as a proof of concept to capture a citywide measure of disparity in the use of frisking, measuring the differential perception of dangerousness based on the perceived race of the subject of the stop. This component builds on this premise and applies the methodology to specific geographies, units of officers, and time periods to track levels of disparity over time and across levels of aggregation.

2.2.1 Initial Considerations and Current Resolution

1. **Balance**. One of the primary considerations in trying to make this approach dynamic over time is balance. How can the availability of good, matched cases be maintained and support the frequency of the EAQ? Certainly, the number of Terry stops cannot and should not be altered to satisfy this model, so SPD's control is limited to the level of aggregation of the method. Before the cadence for both EAQ and this equity component were set to the monthly level, we had suggested drawing matches from a wider timeframe to create higher quality counterfactuals. Over the development of this methodology, it became clear that the model is stable and sufficiently powered from a month's worth of stops. This allows the approach to provide some temporal insights about post-stop disparities without relying on too few stops. We recommend serious consideration before applying this methodology to smaller units of analysis (shorter timeframe or smaller geographic areas).

2. **Number of events**. When using logistic regression to create the propensity scores, there are limitations to the number of events per variable (EPV) that can be employed. Based on the 119 variables used to explain post-stop differences and the intention to divide the pool of potential matches by time and place, we were initially concerned with overfitting the model and violating the EPV assumption. SPD's Bayesian approach (XGBoost) to the computation of propensity scores can relax this assumption and justifies the inclusion of a substantial number of covariates.

3. **Officer behavior**. One of our early concerns with using post-stop outcomes as the proxy measure for racial disparity is the potential chilling effect that measurement itself may have on officer stop and frisk behavior. This sort of de-policing effect has been observed across agencies subject to the consent decree process (Stone et al., 2009; Chanin & Sheats, 2018). Although we have no reason, at this time, to believe that this has occurred in SPD, it is important to monitor as EAQ progresses and becomes a part of officers' daily lives. Whereas we expect that effective communication about how post-stop outcomes will be monitored at an aggregate level may mitigate this effect, SPD has committed to tracking any de-policing effect. Because the model itself relies on a sufficient number of cases for matching, this method would naturally erode if officer drawback occurred at a substantial level.

4. **Staffing**. We have observed that due to staffing availability, and post-pandemic declines in proactivity, the number of qualifying Terry stops is lower than past levels and may affect match quality or the ability to generate consistent results across every month or among certain officer units. We reiterate that there should be a built-in expectation that this metric can only operate with sufficient stops to inform the model and interruptions in this metric may naturally occur. Based on our observations, SPD's workflow continuously monitors the balance of matches and will not relax model parameters for a potentially faulty value if and when these interruptions do occur.

5. **Missing data**. Along the same lines, missing values persist in the data. An initial look at the historical data shows that 12.1% of Terry stop records have missing or unknown data for the race field. This is currently being monitored by a data governance program (DGAL-192). Since these initial reports, SPD has required the race field to be filled out for all filed contacts where Terry stops are reported. Taken together, this is a methodology where upwards of 15% of stops are not directly pertinent to this question of racial disparity, and another approximately 10% may not have data related to the central question of race. These present a baseline of limitations to the data that may confound initial results but should motivate additional efforts to filter out unwanted cases and improve data quality for race and all covariates.

2.2.2 Outstanding Considerations

Non-Terry stops. Based on initial evaluations of the data, 85% of the stops in the model data are Terry stops, with the remaining 15% including probable-cause stops and post-arrest frisks due to an excess of caution in reporting. The inclusion of non-Terry stops does introduce some unwanted noise in the results; however, this is unlikely to change. These estimates are still reasonably accurate, and although there may be future attempts to flag unwanted stops, there is an expectation that over-reporting of stops will still occur and be included in this analysis.

2.2.3 Recommendations & Conclusion

Overall, this methodology is an elegant approach to isolating the "causal" effect of perceived subject race on officer decision making. The propensity score-based approach, which can only theoretically define causality when all known and unknown covariates are controlled for, is an appropriate and sophisticated substitute for a randomized controlled trial, which is not possible here. The use of post-stop outcomes as a measure for the general impact of race on the outcomes of police interactions is appropriate, as these situations are highly discretionary and offer a rich dataset to control for and isolate the effects of race.

That said, the disparity metric associated with this EAQ component is a proxy measure, a relevant avenue to get at the larger question of racial motivation in officer decision making. At no point should the measure of disparity in post-stop outcomes be viewed as a comprehensive measure of racial disparity across police interactions as a whole. To illustrate, this metric does not measure what is potentially the largest source of disparity in Terry stops: the decision whether to stop someone. Gau and Brunson (2010) and Bandes and colleagues (2019) provide support for the importance of the stops themselves being an impactful source for potential disparity regardless of whether a frisk occurs. We understand that future additions to this EAQ component, including risk-adjusted disparity, may expand the scope, but it is important to frame the current view of post-stop outcomes as only part of the complete picture of disparity.

The utility of this metric is contingent on continuity in stop behavior over time and across the unit divisions to which it is applied. If officers conduct fewer stops or proactivity declines to a point that the models cannot compute a summary disparity score, this does not serve as an indicator that racial disparity has been eradicated, but rather that the source of its measurement is no longer available and must be measured in other ways. Ensuring continuity in the number of qualifying events is in

the best interest of this endeavor. This is not to suggest a mandated increase in the frequency of stops, but rather that effective communication to officers and transparency about how this metric will be used may circumvent any pullback behaviors or a breakdown in this measurement.

There are a few ways this component can be expanded. For example, there is value to conducting targeted interviews or focus groups with officers to better understand their decision making when it comes to Terry stops and post-stop frisking. Any additional knowledge, even if qualitative, may improve future modeling and identify relevant covariates. Any additional explanation for discretion will eat into the variability explained by race, reducing any undue disparity attributed to it.

It may be useful to explore other post-stop outcomes beyond frisking to proffer a more comprehensive look at differential burden by race. Comparisons of the existing comparison groups can be used to examine outcomes such as duration of the stop, justification for the frisk, number of officers present, the likelihood of use of force, and frisk outcomes like identification of a weapon or arrest of the subject. These supplementary analyses likely are not feasible as a dynamic measure but may present aggregated measures of these differences as a measure of disparity.

2.3 Summary

We believe that this EAQ metric leverages proven methodologies and existing data collection efforts to provide a reasonable, if limited, measure of racial disparity in officer decision making. We recommend that this component proceed as an integral part of the equity component of EAQ fora, with additional concurrent efforts to improve data quality and, therefore, the accuracy of this measure.

3. Component 2: Location-Based Resource Accountability

3.1 Evaluating the Premise

The primary goal of this EAQ component is to create a measure of the ratio between community need for police in a specific area and the corresponding level of police presence, to flag areas of over- and under-policing. Identification of these areas with service disparity will allow for the investigation into the causes for this mismatch in service and address behavior or planned resource allocation. Using spatial analysis will plot a known concentrations of community demand for service and using AVL data will identify where officers spend time. Assessing the overlap in these measures can demonstrate a proper dosage of police presence but will also allow for the mapping of areas with misalignment. The sourcing of adequate data and the ability to process those data efficiently will determine the appropriateness of this proposed methodology.

To assess overlap, we must measure both community need and officer presence. The proposed measure of community need is derived from the historical volume of calls for service (CFS) to the police, geocoded to its place of origin. The proposed data source for measuring the levels of police presence is the AVL data, which provide an approximation of where officers spend their time. Both data sources are reasonable proxy measures for community need and police response. The use of spatial analytics to determine areas of concordance and disparity in alignment between need and police service is a reasonable solution, as long as careful attention is paid to call criteria used to predict need and the level of geography at which the metric is aggregated. As explored in depth in the pre-implementation report, we find that both the availability of the data and the approach to analysis satisfy the research premise.

3.2 Validating the Methodology

The use of citizen CFS to approximate need and the use of AVL data to measure police presence are both well-established methodologies. The concentration of crime in a few high-volume locations is an accepted way to define and direct police patrol to these areas of high need (Sherman & Weisburd, 1995; Weisburd, 2015). Seattle's approach—leveraging predictive models of need based on past concentrations—is an extension of this logic, whose value corresponds to the quality of the call inclusion parameters. Likewise, AVL is a commonly used metric for where officers spend time (Weisburd et al., 2015; Wu et al., 2022; Telep et al., 2014).

The innovation of the current approach is examining the geographic relationship between these two metrics. Analogous studies focused on proactive policing (Wu & Lum, 2017) provide a logical foundation for exploring the overlap of police activity at a higher level of specificity. Seattle's current application of AVL and CFS analysis will largely follow initial efforts to explore these concepts in relation to the community's demand for enforcement (Atherley et al., 2022). This work was originally developed as part of a routine SPD research project that applied these insights after the development work was complete. Pending successful hurdling of technological limitations, this component is a feasible measure of the spatial and temporal overlap between demand for service and police adherence to providing that service. The primary contention with this approach is the balance between masking variability between individual streets using density-based clustering to identify larger areas, and the operational need to identify specific areas of interest that may not conform to a traditional street-based approach.

3.2.1 Initial Considerations and Current Resolution

1. **Anticipated need**. We initially pointed out limitations to using a cross-sectional approach in defining public demand for police service. There is some evidence of temporal stability in crime hot spots, but they are often classified into increasing or decreasing CFS trajectories over time. SPD is adopting a dynamic predictive approach, where the call time and location are used to define anticipated need in each of the prediction zones. What the refresh rate for service need may be is unclear, but the ability to iterate on both demand and officer presence is essential, as EAQ carries on for any extended period of time.

2. **Perceptions**. There is an outstanding question of whether community members' perceptions of over- and under-policing match the data. Although the current measure is a primarily a practical measure of accountability, this EAQ component may present a future opportunity to assess equity in service delivery compared to perceived demand and the demand articulated by community perceptions.

3. **Definition of need**. Careful attention must be paid to how need is defined within a community, specifically as it relates to sequestering resident-initiated and officer-initiated calls. Officer-initiated, or on-view, calls may be more indicative of where presence is targeted than of community need. Using existing officer activity to define need may introduce a self-justifying feedback loop, where need is defined as where officers already spend time and

initiate calls. For this reason, we suggest—and SPD agrees—that need is indicated by dispatched CFS to the police.

4. **In-transit data concerns**. Initially, we raised concerns about how AVL transmission while using thoroughfares or traffic corridors may indicate a concentration of officer presence without community need tied to those locations. Whereas it is true there will be no differentiation in AVL data pings between on-scene and in-transit status, newer features of the methodology mitigate the problem presented here. Foremost, the change in geographic clustering methods to a density-based approach (DBScan) results in 796 zones, compared with the nearly 2,500 zones previously predicted using affinity propagation. Operating at higher levels of geographic aggregation means that in-transit data points are likely to be dispersed across zones and be not problematic.

5. **Service levels at locations**. In this methodology, officer presence is assumed to be in service of community need. However, some locations are likely convenient places for administrative tasks, report writing, or meal breaks. As such, AVL location data during these times should not necessarily contribute to a measure of service-levels, although these concentrations should not and will not be disregarded from the overall analysis. These known concentrations may represent potential operational security concerns (ambush risk) and should be known. As part of the EAQ process, these individualized locations are identified and explained in the context of over-policing. Because this is an iterative process, known hot spots can be annotated and subsequently filtered out of EAQ conversations, once they have been addressed.

3.2.2 Outstanding Considerations

1. **Officer behavior vs. presence**. The differentiation between the time spent in an area and the activities conducted while in that area is absent from this EAQ metric. It may be important to capture the activities of the officers beyond when and where the AVL pings their locations. There is a functional difference between an officer driving from point A to point B through a neighborhood and an officer engaging in a 15-minute directed patrol on that block face (Nagin et al., 2015; Koper, 1995). Furthermore, a directed patrol where an officer spends time in the car is different from both community policing efforts or proactive law enforcement, which contribute differently to perceptions of over- and under-policing.

The data here are necessarily limited to officer locations. However, as this EAQ component continues to develop, considering the effects of officer behavior beyond mere presence will be important.

2. **Street segment variability**. This EAQ metric balances having enough granularity to focus on specific locations, while also maintaining a high enough geographic aggregation for predictions to be valid. We believe that SPD's intended approach strikes this balance, but we also contend that the larger polygonal divisions of the city will mask variability that may be relevant to identifying areas of over- or under-policing. The heterogeneity of crime and community need between street segments within a community is well documented (Weisburd et al., 2004; Steenbeek & Weisburd, 2016). First observed in Seattle, streets next to each other, even in places classified as "bad neighborhoods," can have very different needs for police presence, based on the heterogenous distribution of crime at the street level. The current approach, generalizing both need and presence at a meso-geographic level, can miss variability in both within the defined areas. This criticism is not intended to discount the current method, but suggest additional levels of analysis for the future, facilitated by the point level data collection of CFS and AVL pings.

3.2.3 Recommendations & Conclusion

Overall, this methodology satisfies its intended goal of creating a measure of resident need and police service. This success is contingent on functionable AVL data collection and management, and the accurate prediction of need using continuously updated CFS data. The identification of misalignment between these two spatial data layers is a creative solution to identify service disparity. These data sources are both appropriate and available, and the metric itself relies on well-established spatial statistics to identify a novel need.

Beyond its stated goals, there is potential for a positive, unintended consequence of this component. Although they note the importance of the role of leadership and organizational history, de Brito and Ariel (2017) find that the act of monitoring patrol locations can increase fidelity to assigned patrols. However, it is important to define the scope of what this metric can really say. We contend that the residuals indicating over- or under-policing compared to need should serve as conversation starters in the EAQ process, and not a definitive measure of over-policing as experienced by the community. What this metric is best positioned to do is highlight the service areas with the greatest

disparities and serve as an inflection point for considering why that disparity exists and whether any intervention is necessary.

Owed to the meso-level of geographical abstraction and the expectation that officer presence can legitimately go beyond the immediate needs of citizen crime calls, this outcome measure is not necessarily a measure of true inequity but a signal of where it might be found with further investigation. For these reasons, this EAQ metric should be framed as a useful operational tool for better managing police resources and justifying existing police presence, rather than an academic measure of how and when the public experiences these service disparities.

There are a few ways this component can be expanded. Currently, this metric relies on AVL data as a measure for police service in a community; it is the best and most accessible metric available at the timescale required for continuous monitoring. However, AVL can capture more than the strict definition of police service. Likely dispersed across the city, AVL pings during transit, meal breaks, or other administrative tasks that all contribute to our understanding of where police are engaging throughout the city. Referring to the importance of officer behavior as much as presence, as EAQ progresses, the measurement of police service might be refined to include only those officer activities that may contribute to perceptions of over- or under-policing.

Because the source data for both AVL and citizen need (call location) are at the point level, there is a rich potential to explore these concepts at a lower level of geographic aggregation. Although these efforts may not be appropriate for continuous monitoring as part of the EAQ, examination at the street segment level may help to unshroud the masked variability discussed earlier. At this microgeographic level, conversations about streets with service disparities can become a lot more specific. We believe this is worth exploring in conjunction with the current planned measure.

3.3 Summary

This EAQ component leverages well-established and accessible data sources as a reasonable measure for where police are spending time and where they ought to spend time. Due to the discussed limitations to geographic specificity and the inability to distinguish what officers are doing, we do not view this metric as an academic or definitive measure of over-policing. However, the metric does serve as a valid approach to initiating conversations about why observed disparities may occur and iteratively improve on service equity from there. We recommend this component proceed as a valuable part of the EAQ forum

4. Component 3: Officer Service Quality Assessment

The format of Component 3 will differ slightly due to the recent transition from automated body camera transcription to targeted surveys of citizens with recent police contact. This new component was not featured in the pre-implementation report, and this constitutes a first iteration of evaluation and recommendation.

4.1 Evaluating the Premise

The original stated goal of this EAQ component was to create an engaging measure of the quality of interactions between officers and community members. The use of body-worn video and classification models was the most robust methodology to achieve both the frequency and scope of the intended metric. With this methodology no longer in use, the question becomes the suitability of targeted community surveys to satisfy the EAQ goal. The appropriateness of this methodology is predicated on the belief that the collected data are both attainable and believable.

In January 2023, the department implemented a continuous measurement satisfaction survey, administered by an automated platform also used to update those accessing police service. After a community member calls 911, the system sends a set of automated messages (text and/or email) to the contact information they provided. This message confirms their request for service, provides reference information, and some limited instructions preparing them for the response (e.g., documents and materials to have available for an auto theft report), if applicable. After the officer completes service (clears the call), additional automated messages are sent asking if the community member would like to participate in a satisfaction survey. Questions about service satisfaction are presented using the Net Promoter format, whereby the community member is asked whether they would refer a friend or family member dealing with a similar issue to request service from the SPD. Additionally, change in fear of crime questions are asked. The subject is asked if their specific interaction increased or decreased their fear of crime during the day and at night, separately. Some demographic and use type (e.g., resident of the city, works in the city) and unstructured free text response are included.

Satisfaction questions are relative to the resource and phase of service the community member recently interacted with. During the initial response, the Net Promoter question is asked relative to the person the community member spoke with on the phone, the officer, and the department as a whole. If the subject is listed as a victim in a police report, after that report processes through the

Records Management System (RMS), up to 12 hours after the response, an additional victim satisfaction survey is processed. If the report results in a follow-up investigation by a detective, additional satisfaction questions are asked using the same automated process. Finally, once the case is closed (e.g., inactivated, referred for prosecution, declined by the prosecuting attorney's office), a final survey is initiated.

The Net Promoter format was selected to control survey effects and provide immediate comparability across analogue industries. As indicated previously, it is assumed the responses are biased toward those with a motivation to respond. Responses are assumed to reflect those who are extremely satisfied and extremely dissatisfied with the service they received. Given the potential to reactivate the trauma of a person who was recently the victim of a violent crime, all violent crimes against person are removed from the automated survey process. This control may eliminate an overly positive response from someone who is grateful for having had their physical safety protected, directly. The residual emotional effect (midbrain) of a highly stressful, frustrated, or otherwise victimized feeling is moderated by the Net Promoter question format. This question asks the respondent to consider whether, based on part or the totality of their experience, they might recommend a friend or family member take similar action to access services. This referential consideration deploys some additional cognitive processing, reflecting a value judgment made about an external object (a friend or family member) and is commonly employed in customer satisfaction where emotional or impression managed responses are a risk. In addition to allowing for a manageable dimension (complexity and scope of the instrument), increasing response and completion rates,[1] the Net Promoter model does not require a new scale be validated (e.g., test retest reliability, interrater reliability) and provides immediate comparability across industries.

The SPD intends to track movement, as well as cross-industry comparables, for engagement of this metric. Although policing generally suffers from a lack of comparability, the highly emotional and selection-biased nature of satisfaction responses compounds the problem. The cross-industry comparability of this metric allows the SPD and stakeholders to contextualize satisfaction scores in a meaningful way. Additionally, as is the case with disparity measures under the Equity metric (above), trends and patterns provide actionable insights. Whereas a good equity metric can be said to be *as low as reasonably achievable* (ALARA), a common strategy for managing key performance indicators in

[1] Initial operation suggests a sustained response rate of 20% over the first 2 months of operation.

safety and risk, the "good" of a quality metric can be said to be *as high as reasonably achievable* (AHARA). Identifying opportunities to increase and/or optimize quality metrics is the goal of the Quality component of EAQ and is achieved through the use of this method. In this way, it is the movement and relative context (outlier) of the quality measurement that is actionable; the effect of selection bias is effectively mitigated by its intended form of engagement.

Historically, community surveys are substantively robust and infrequent due to the cost and effort of implementation. However, SPD's leveraging of SPIDR Tech automates the dissemination and collection of surveys and supports the premise that interaction data are widely attainable. There is an inherent loss in specificity of the metric because the data source shifts from an objective record of the interaction to the post facto perception of the surveyed resident. The SPD has given significant consideration to the selection bias inherent in this approach. As this is a convenience sample, and participation is voluntary, without incentive, it is assumed that respondents are motivated to respond. Motivation is both positive and negative: Respondents may be motivated by either an extremely high or extremely low subjective perspective on the service delivered.

However, there is no reason to doubt that the limited scope of questions asked in these surveys is believable. In theory, the use of frequent feedback can provide reasonable estimates for any measure included in the survey, although this is contingent on acquiring a sufficient volume of responses, because participation is neither automatic nor compulsory. In reality, it is important to consider whether the subset of people who do respond and their perceptions are representative of the interactions overall; these specific questions are considered below.

The logic behind this EAQ component does, in theory, present a reasonable data collection and analysis protocol that can address narrow aspects of interaction quality.

4.2 Validating the Methodology

The use of community surveys to evaluate perceptions of the police and police interactions is a well-established methodology and is becoming increasingly common (Rosenbaum et al., 2017; Merenda et al., 2021). Traditionally, these surveys are cross-sectional and provide estimates at the population level. Here, Seattle is leveraging SPIDR Tech to make these surveys more targeted and frequent following any qualifying interaction with the police, enabling a dynamic estimate of perceptions of interaction quality over time. In theory, this works to establish a baseline in the population and then to apply repeated measures to the same population to detect changes.

These approaches—using immediate post-interaction surveys—have not been academically or scientifically evaluated, although they are in use throughout public safety offices and police agencies across the country. Because of the lack of research in this area, we treat the methodology as a logical extension of the larger community surveys that are validated and used to address these same questions.

One novel expansion of this approach in Seattle is the use of Net Promoter Scores (NPS). This evaluation of services, based on whether or not one would recommend this service to a friend or relative in the same situation, is common in marketing research (Fisher & Kordupleski, 2019), but has also been expanding to the medical fields (Krol et al., 2015) and public sector agencies (Luoma-aho et al., 2021). In examining the survey content specific to Seattle, we found the potential that focusing on recommendation of services may be (1) tied to either global attitudes about the police beyond the scope of the most recent interaction, or (2) driven more by outcomes than the process and treatment during the interaction (Tankebe, 2013). Given the restrictions on more direct measures of interaction quality, we see potential in this method but raise the following initial considerations as the measurement is further integrated in the EAQ process.

4.2.1 Considerations

1. **Response rate**. It is important to consider the response rate, when extrapolating survey responses into a global measure of citizen satisfaction and interaction quality. Response rate measures how many responses were received out of how many could have been. SPIDR Tech self-reports an average response rate to their post-interaction surveys at about 12.1%. According to initial data, Seattle's survey response rate is between 20% and 25%, which is better than the SPIDR Tech baseline and generally considered within range for NPS scores in other fields. The importance of response rate is contingent on the minimum viable number of respondents and the size of the effect to detect. Distributing surveys to a majority of police interactions daily still results in a large population of surveys from which to draw results. However, the risks of extrapolation increase while drawing on a smaller percentage of the overall population. Here, we risk the assumption that the 20% who respond are behaviorally and substantively the same as the 80% who do not. This raises the next consideration of selection bias.

2. **Selection bias** occurs when the responses received do not represent the population generally. There are two opportunities for selection bias in the administration of these surveys. First, the post-interaction surveys are not universally applied. Whereas the goal of this EAQ metric is to offer a proxy measure for interactions generally, the survey appears to be limited to those who willingly engaged with police in the first place. This excludes victims of violent crime, subjects of proactive police enforcement, traffic stops, and arrestees. Exclusion of this segment of the population ignores the measures of interaction quality in those scenarios where it is potentially most important. The second potential for selection bias comes from the respondent in their decision whether to respond. As indicated by the response rate, there is variability in whether people complete the survey, which poses the question of why these discrepancies exist. As with other opportunities for feedback, only those with the best and worst experiences may be willing to take the time to offer praise or criticism. Understanding and accounting for selection bias is a huge hurdle for the believability of this metric.

3. **Call type**. The cadence of these surveys allows for a highly dynamic measure that can be examined at a higher frequency than the other components, at a weekly or even daily level, to identify or explain outliers. The EAQ metrics are not designed to use individual officers as the unit of analysis; however, this metric is well set up to disaggregate the overall scores by beat or unit, which can serve as part of the incentive structure. One additional unit of analysis may be the call nature related to the survey. Although disaggregation to the call type level may not be possible or supported by the call volume at the same time as resolution, organizing interaction quality by type of call would provide some insights about which situations (for both officers and subjects) may be driving low- or high-quality ratings. This has the operational benefit of targeting additional trainings or interventions to improve ratings and NPS.

4. **Net Promoter Score**. The primary outcome measure in the survey is the NPS focused on whether these services would be recommended to someone else in a similar situation. As discussed, this may be driven by distributive justice outcomes rather than procedural fairness or professionalism. However, it does carry the unique benefit of facilitating cross-industry comparisons, which can serve as an anchoring point for understanding of and conversations about overall performance. The included measure of interaction quality is an open text field

that is not conducive to identifying trends in officer behavior, demeanor, or performance. SPIDR Tech has survey item templates directly related to officer courteousness that more directly mirror the previous attempted measure of quality. This is worth considering as an addition to the current survey.

5. **Perceptions**. In interpreting the survey results and creating the overall metric, it may be difficult to separate out responses driven by the latest interaction with the police, and engrained perceptions due to a history of direct and vicarious interactions with the police. There is evidence that perceptions of the police generally can be affected by recent direct experiences, neighborhood context, vicarious experiences shared by family, and long-held intergenerational beliefs (Harris & Jones, 2020; Wolfe et al., 2017; Fine et al., 2022). In this context, a high NPS may be due to a positive outcome from their most recent interaction, a professional experience high in procedural justice, or a stable belief in the legitimacy of police that manifests regardless of the immediate situation. Whether the reasoning for the score may be teased out is unclear, but adding questions explicitly about the context and outcomes of interest for the most recent interaction may provide additional insight.

6. **Demographics**. Previous evaluations of community perceptions of police satisfaction find that demographics such as age, education, race, and fear of crime can explain some of their ratings, beyond direct experience with the police (Haberman et al., 2016; Weitzer & Tuch, 2005). Seattle's post-interaction surveys contain many demographic questions and, contingent on the completeness of that data, these may be used to explain any observed differences in NPS or satisfaction beyond the context of the most recent interaction.

4.2.2 Recommendations & Conclusion

Because this EAQ component was not evaluated in the same iterative way as the others, the initial considerations comprise our recommendations for organizing and moving forward with this metric. Overall, we believe that the post-interaction survey satisfies the conditions for generating a narrow measure of interaction quality. Compared to the original transcription and analysis of body-worn video, reliance on citizen perceptions introduces an additional degree of subjectivity both in terms of content and the decision to participate. Although this method is an acceptable approach for following up on officer interaction quality, there are some critical changes to the actual content of the surveys that would improve this metric.

As it stands, the content is focused on global measures of police performance and approval. As discussed, these opinions can be colored with experiences far beyond the interaction in question. Either expanding or replacing the survey content to explicitly measure officer professionalism, sentiment, and procedural justice would be valuable additions in line with the goals of this EAQ component to ensure quality across a range of interactions. In extrapolating the results of these surveys, accounting for the limitations of this methodology—including selection bias, exclusion of certain types of interactions, and content that may go beyond the most recent encounter—will be essential. As the methodology develops to consider these recommendations, it may serve as a core quality metric for the EAQ program.

5. The Seattle Crime Harm Index

Rather than focus entirely on raw crime counts, practitioners and researchers have begun to examine the use of harm indexes as a way of analyzing crime. Crime counts do not distinguish between the total number of property thefts versus a robbery.

An index attempts to create consistency across disparate variables. The intent of harm indexes is to create a numerical value for crime that equalizes the type of crime by the amount of harm it generates. The first harm index created used court records as an indicator for determining harm (Ignatans & Pease, 2015). Researchers built on this approach, adding in metrics like traffic accidents and drug offenses (Ratcliffe, 2015), while others focused on victim harm (Greenfield & Paoli, 2013). However, Sherman (2013) was the first to create a crime harm index (CHI) based on sentencing guidelines for first offenses (Sherman 2007, 2013; Sherman et al., 2016).

The SPD wanted to incorporate this approach into their crime analysis process and constructed their own harm index based on Washington State offense codes. SPD chose to follow Sherman's original (2013) Cambridge CHI method and used sentencing data to create their CHI.

The creation of the Seattle Crime Harm Index (SCHI) will allow SPD to compare crime statistics by crime type without losing the variance of the harm some crime causes compared to others. CHI levels the variance between high-volume/low-harm and low-volume/high-harm crime. Sherman (2013) outlined the following steps for creating the Cambridge CHI, using the median number of prison days to calculate crime harm:

1. Count the number of crimes of each type.
2. Multiply the count for each type by the median number of prison days recommended for crimes of that type by first offenders.
3. Call the product of that multiplication (crime count for a crime type X median days in prison) the harm subtotal (HST) of days of prison for that offense type.
4. Repeat steps 1, 2, and 3 for every type of crime recorded for the area or person.
5. Sum up all HSTs to yield the total crime harm (TCH).

SPD emulated the Cambridge CHI steps to develop the SCHI. SPD began the construction of the SCHI by requesting and receiving offense-level data from the Washington State Center for Court

Research, Administrative Office of the Courts. SPD requested data from 2008 to Nov. 21, 2021, to include outcomes (guilty/not guilty) and the resulting sentence for first offenses. SPD calculated the average length in days of sentences for first offenses without sentencing "enhancements" where the verdict was guilty. This became the SCHI value.

Where SPD could not calculate an average for the specific Revised Code of Washington (RCW), SPD consulted the Washington State Adult Sentencing Guidelines Manual. SPD took an average of the highest and lowest of the sentencing range (without "enhancements") to determine the index value. Monetary fines were converted to the index value by taking the dollar value and dividing it by the minimum wage for the City ($13.50 in 2020) to determine the number of hours. The result was divided by 8 to create an equivalent to the sentencing days. Finally, to produce a severity score, SPD matched data from the Incident-Offense Data Source and assigned a severity score based on the type of aid response and the Seattle CHI score.

However, SPD was still awaiting information about misdemeanor sentences from the municipal court as of March 2023. Prior to 2020, SPD had a well-established relationship with a data analyst with the municipal court. Unfortunately, that resource left the court; since then reconnecting with municipal court has been difficult. Like all City resources, the court has an overload of public requests and is still recuperating from staffing turnover issues.

The primary goal of SPD was to incorporate the SCHI values into the data warehouse (DAP) Incident-Offense data source. When the SCHI is complete and included in the DAP, it will allow SPD and their research partners to study the concepts of harm in policing and develop alternative deployment strategies. Currently, in the prototype and proof of concept, SPD is only utilizing the SCHI component as an additional identifier in the murder/homicide cases as part of the match verification. SPD expects to have a fully functioning model once they can retrieve sentencing data from the municipal court.

References

Component 1:

Bandes, S. A., Pryor, M., Kerrison, E., & Goff, P. (2019, February 6). The mismeasure of Terry stops: Assessing the psychological and emotional harms of stop and frisk to individuals and communities. *Behavioral Sciences & the Law*, *37*(2), 176–194.

Chanin, J., & Sheats, B. (2018). Depolicing as dissent shirking: Examining the effects of pattern or practice misconduct reform on police behavior. *Criminal Justice Review*, *43*(2), 105-126. https://doi.org/10.1177/0734016817704696

Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics*, *84*(1), 151–161. https://doi.org/10.1162/003465302317331982

Gau, J. M., & Brunson, R. K. (2010). Procedural justice and order maintenance policing: A study of inner-city young men's perceptions of police legitimacy. *Justice Quarterly*, *27*(2), 255-279. https://doi.org/10.1080/07418820902763889

Govindasamy, P. (2016). *A comparison between propensity score matching, weighting, and stratification in multiple treatment groups: A simulation study* [Doctoral dissertation]. University of Denver. https://digitalcommons.du.edu/etd/1173/

King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political Analysis*, *27*(4), 435–454. https://doi.org/10.1017/pan.2019.11

Levchak, P. J. (2021). Stop-and-frisk in New York City: Estimating racial disparities in post-stop outcomes. *Journal of Criminal Justice*, *73*(March-April), 101784. https://doi.org/10.1016/j.jcrimjus.2021.101784

Olmos, A., & Govindasamy, P. (2015). A practical guide for using propensity score weighting in R. *Practical Assessment, Research, and Evaluation*, *20*(13), 1–8. http://pareonline.net/getvn.asp?v=20&n=13

Rosenbaum, P. R., & Rubin, D. B. (1983, April). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Seattle Police Department (SPD). (2019). *Disparity Review – Part 1*. Case 2:12-cv-01282-JLR Document 554-1.

Stone, C., Foglesong, T. S., & Cole, C. M. (2009). *Policing Los Angeles under a consent degree: The dynamics of change at the LAPD*. Cambridge, MA: Program in Criminal Justice Policy and Management, Harvard Kennedy School.

Component 2:

Atherley, L. T., Hickman, M. J., Parkin, W. S., & Helgott, J. B. (2022). Measurement of potential over- and under-policing in communities. *Policing: A Journal of Policy and Practice*, 16(3), 443-461. https://doi.org/10.1093/police/paac025

de Brito, C., & Ariel, B. (2017). Does tracking and feedback boost patrol time in hot spots? Two tests. *Cambridge Journal of Evidence-Based Policing*, *1*, 244–262. https://doi.org/10.1007/s41887-017-0018-7

Koper, C. S. (1995). Just enough police presence: Reducing crime and disorderly behavior by optimizing patrol time in crime hot spots (NCJ Number: 167664). *Justice Quarterly*, *12*(4), 649–672. https://doi.org/10.1080/07418829500096231

Nagin, D. S., Solow, R. M., & Lum, C. (2015). Deterrence, criminal opportunities, and police. *Criminology*, *53*(1), 74–100. https://doi.org/10.1111/1745-9125.12057

Sherman, L. W., & Weisburd, D. (1995). General deterrent effects of police patrol in crime "hot spots": A randomized, controlled trial (NCJ Number: 167663). *Justice Quarterly*, *12*(4), 625–648. https://www.ojp.gov/ncjrs/virtual-library/abstracts/general-deterrent-effects-police-patrol-crime-hot-spots-randomized

Steenbeek, W., & Weisburd, D. (2016). Where the action is in crime? An examination of variability of crime across different spatial units in The Hague, 2001–2009. *Journal of Quantitative Criminology*, *32*, 449–469. https://doi.org/10.1007/s10940-015-9276-3

Telep, C. W., Mitchell, R. J., & Weisburd, D. (2014). How much time should the police spend at crime hot spots? Answers from a police agency directed randomized field trial in Sacramento, California (NCJ Number: 248011). *Justice Quarterly*, *31*(5), 905–933. https://doi.org/10.1080/07418825.2012.710645

Weisburd, D. (2015). The law of crime concentration and the criminology of place. *Criminology*, *53*(2), 133–157. https://doi.org/10.1111/1745-9125.12070

Weisburd, D., Bushway, S., Lum, C., & Yang, S. M. (2004). Trajectories of crime at places: A longitudinal study of street segments in the city of Seattle. *Criminology, 42*(2), 283–322. https://doi.org/10.1111/j.1745-9125.2004.tb00521.x

Weisburd, D., Groff, E. R., Jones, G., Cave, B., Amendola, K. L., Yang, S. M., & Emison, R. F. (2015). The Dallas patrol management experiment: Can AVL technologies be used to harness unallocated patrol time for crime prevention?. *Journal of Experimental Criminology*, *11*, 367–391. https://doi.org/10.1007/s11292-015-9234-y

Wu, X., & Lum, C. (2017). Measuring the spatial and temporal patterns of police proactivity. *Journal of Quantitative Criminology*, *33*, 915–934. https://doi.org/10.1007/s10940-016-9318-5

Wu, X., Koper, C., & Lum, C. (2022). Measuring the impacts of everyday police proactive activities: Tackling the endogeneity problem. *Journal of Quantitative Criminology*, *38*, 242–262. https://doi.org/10.1007/s10940-021-09496-8

Component 3:

Fine, A. D., Padilla, K. E., & Tom, K. E. (2022). Police legitimacy: Identifying developmental trends and whether youths' perceptions can be changed. *Journal of Experimental Criminology*, *18*, 67–87. https://doi.org/10.1007/s11292-020-09438-7

Fisher, N. I., & Kordupleski, R. E. (2019). Good and bad market research: A critical review of Net Promoter Score. *Applied Stochastic Models in Business and Industry*, *35*(1), 138–151. https://doi.org/10.1002/asmb.2417

Haberman, C. P., Groff, E. R., Ratcliffe, J. H., & Sorg, E. T. (2016). Satisfaction with police in violent crime hot spots: Using community surveys as a guide for selecting hot spots policing tactics (NCJ Number: 249608). *Crime & Delinquency*, *62*(4), 525–557. https://doi.org/10.1177/0011128713516840

Harris, J. W., & Jones, M. S. (2020, March-April). Shaping youths' perceptions and attitudes toward the police: Differences in direct and vicarious encounters with police. *Journal of Criminal Justice*, *67*, Article 101674. https://doi.org/10.1016/j.jcrimjus.2020.101674

Krol, M. W., de Boer, D., Delnoij, D. M., & Rademakers, J. J. (2015). The Net Promoter Score–an asset to patient experience surveys? *Health Expectations*, *18*(6), 3099–3109. https://doi.org/10.1111/hex.12297

Luoma-aho, V., Canel, M. J., & Hakola, J. (2021). Public sector reputation and netpromoter score. *International Review on Public and Nonprofit Marketing*, *18*, 419–446. https://doi.org/10.1007/s12208-021-00280-9

Merenda, F., Trent, J., Rinke, C. R., & Buchanan, M. (2021). Understanding citizen satisfaction with the police: Results from a community survey. *Police Practice and Research*, *22*(1), 692–710. https://doi.org/10.1080/15614263.2020.1772784

Rosenbaum, D. P., Maskaly, J., Lawrence, D. S., Escamilla, J. H., Enciso, G., Christoff, T. E., & Posick, C. (2017). The police-community interaction survey: Measuring police performance in new ways. *Policing: An International Journal of Police Strategies & Management*, *40*(1), 112–127. https://doi.org/10.1108/PIJPSM-07-2016-0119

Tankebe, J. (2013). Viewing things differently: The dimensions of public perceptions of police legitimacy. *Criminology*, *51*(1), 103–135. https://doi.org/10.1111/j.1745-9125.2012.00291.x

Weitzer, R., & Tuch, S. A. (2005). Determinants of public satisfaction with the police. *Police Quarterly*, *8*(3), 279–297. https://doi.org/10.1177/1098611104271106

Wolfe, S. E., McLean, K., & Pratt, T. C. (2017). I learned it by watching you: Legal socialization and the intergenerational transmission of legitimacy attitudes. *British Journal of Criminology*, *57*(5), 1123–1143. https://doi.org/10.1093/bjc/azw038

Crime Harm Index

Greenfield, V. A., & Paoli, L. (2013, September). A framework to assess the harms of crimes. *British Journal of Criminology*, *53*(5), 864–886. https://doi.org/10.1093/bjc/azt018

Ignatans, D., & Pease, K. (2015). Taking crime seriously: playing the weighting game. *Policing, 10*(3), 184–193. https://doi.org/10.1093/police/pav029

Ratcliffe, J. H. (2015, June). Towards an index for harm-focused policing. *Policing, 9*(2), 164–182. https://doi.org/10.1093/police/pau032

Sherman, L. (2013, May 20). The rise of evidence-based policing: targeting, testing, and tracking. *Crime and Justice, 42*, 377–343. https://doi.org/10.1086/670819

Sherman, L. (2007, November 9). The power few: experimental criminology and the reduction of harm. *Journal of Experimental Criminology, 3*, 299–321. https://doi.org/10.1007/s11292-007-9044-y

Sherman, L., Neyroud, P. W., & Neyroud, E. (2016, April 3). The Cambridge Crime Harm Index: measuring total harm from crime based on sentencing guidelines. *Policing, 10*(3), 171–183. https://doi.org/10.1093/police/paw003

ATTACHMENT 2

Article

Measurement of Potential Over- and Under-policing in Communities

Loren T. Atherley*, Matthew J. Hickman†, William S. Parkin‡ and Jacqueline B. Helfgott §

Abstract Over- and under-policing of neighbourhoods can undermine public trust and confidence in the police as well as the broader justice process. This study reports on attempts to operationalize and test a spatial indicator of potential over- and under-policing, where over-policing is defined as a level of police presence at a particular location that is greater-than-expected, given the level of public demand for police services, current police enforcement strategy, and community preference regarding police activity. Automated Vehicle Locator data and Computer-aided dispatch logs from the Seattle Police Department, as well as data drawn from community-based surveys, are modelled using a Geographic Information System. The model uses 2-week data windows to provide timely and actionable information that can be rendered for decision makers in a CompStat style accountability and management forum. Such an approach has potential utility for police management, as well as for community engagement and reform efforts aimed at addressing the problem of over-policing.

Introduction

Individual police behaviour is often the subject of intense scrutiny in the wake of high-profile killings of Black, Indigenous, and People of Colour community members. However, police management and the systems by which police leadership exercises control are critical to these outcomes. Disparate over- and under-policing of communities can undermine public trust and confidence across the criminal justice system (Perry, 2006; Hough, 2012; Goldsmith and Harris, 2012;). Traditional approaches to patrol resource management rely on the autonomy and discretion of the officer. Much in the way that machine learning can inherit bias from a training dataset, particularly in the criminal justice system (Yapo and Weiss, 2018), human experiences colour perceptions of reality, and discretionary behaviours are especially subject to this influence. Although some have expressed concern about the extent to which these

*Loren T. Atherley, Seattle Police Department, Seattle, WA, USA

†Matthew J. Hickman, Department of Criminal Justice, Criminology & Forensics, Seattle University, Seattle, WA, USA. E-mail: hickmanm@seattleu.edu

‡William S. Parkin, Department of Criminal Justice, Criminology & Forensics, Seattle University, Seattle, WA, USA

§Jacqueline B. Helfgott, Department of Criminal Justice, Criminology & Forensics, Seattle University, Seattle, WA, USA

Advance Access publication: 11 April 2022
Policing, Volume 16, Number 3, pp. 443–461
doi: https://doi.org/10.1093/police/paac025

limitations can be mitigated (Lum, 2017), awareness of bias or biasing effects are thought to be effective. A more directive approach to patrol deployment and problem solving can mitigate some of these effects. In addition, the analysis of data depicting where police spend their discretionary time is an asset for police managers.

Understanding police patrol behaviours is an important first step to contextualizing community concerns around over- and under-policing. The amount and/or type of police services being provided at a particular location is largely a function of the public demand for police services (e.g. calls for service originating from the 911 system), as well as existing police enforcement strategies for that location (e.g. directed patrol activity and problem-oriented policing). There is also a degree of community preference for police enforcement activity as well as baseline crime levels and tolerance for deviance that may manifest in actual police behaviour, or the 'vigour' of response (Klinger, 1997). These concepts (demand, strategy, and preference) help us to define the expected level of police presence at a particular location, with which one might judge whether an appropriate amount of policing is occurring. Absent the deliberate action of officers (such as organized reductions in service due to labour disputes or other causes) or actual lack of police capacity (which refers to reduction in police services as a result of lack of resources), what remains might then be termed 'under-policing': A lower-than-expected level of police presence at a particular location, given the level of public demand for police services, current police enforcement strategy for the location, and community preference regarding police activity at that location. 'Over-policing' is then the opposite condition: A greater-than-expected level of police presence at a particular location, given the level of public demand for police services, current police enforcement strategy for the location, and community preference regarding police activity at that location.

Over-policing is a frequently heard complaint within some neighbourhoods and it is generally thought to have the greatest potential to undermine public trust and confidence in the police. However, there is limited research that directly examines the relationship between over- or under-policing and other criminological constructs such as hot spots policing and police legitimacy. This lack of research is concerning because of the potentially disproportionate impact on marginalized populations (such as those experiencing homelessness, and persons with mental illness) and disadvantaged communities. For example, to examine over- and under-policing, Boehme *et al.* (2020) operationalized over-policing based on respondent perceptions of excessive use of force in their neighbourhood and under-policing as a scale composed of multiple question responses (e.g. 'How much of a problem is the police not patrolling area or responding to calls from area?', 'Police in neighborhood are responsive to local issues'). These researchers found that persons of colour, to varying degrees, were more likely than white persons to perceive both over- and under-policing as an issue in their neighbourhood. This is consistent with research that has examined the impact of over-policing in indigenous communities (Perry, 2006; O'Brien, 2021) and among other minority communities (Ben-Porat and Yuval, 2012). To be sure, perceptions of over- and under-policing are complicated and the effects of additional police presence on crime are complex and may differ across racial groups with disproportionate burdens, but also disproportionate benefits (Chalfin *et al.*, 2020).

Measuring the dosage of policing in hot spots has long been a subject of interest (Koper, 1995), and our field is beginning to develop methods for estimating treatment fidelity and dosage in micro-locations, as well as making managerial decisions about resource allocation, using Global Positioning System (GPS) tracking devices including radios and Automated Vehicle Locator (AVL) data (Weisburd, 2013, 2016, 2021; Telep *et al.*,

2014; Wain and Ariel, 2014; Kochel *et al.*, 2015; Weisburd *et al.*, 2015; Gibson *et al.*, 2017; Mitchell, 2017; Blanes i Vidal and Matrobuoni, 2018; DeAngelo *et al.*, 2020). In particular, DeAngelo *et al.* (2020) and Weisburd (2013, 2016, 2021) have demonstrated the utility of AVL data as a general indicator of police presence for exploring response time, car accidents and injury, and crime preventative effects of police patrol. For example, Weisburd (2021) aggregated AVL data to the hourly level within beats in Dallas, TX, and used a novel instrumental variable (assignment of patrol vehicles to calls outside their assigned beats) to study the effect of police presence on crime, finding that a 10% decrease in police presence resulted in a 7% increase in crime.

Following this groundbreaking body of work, we offer a somewhat similar approach for identifying potential over-policing that puts police departments and the communities served in a better position to address strained police–community relations. While these earlier efforts are focused primarily on the effect of police presence on crime, our focus is on how police presence might affect community perceptions of police. Thus, the presence of a police vehicle may contribute a deterrent effect on crime, but at what potential cost to community perceptions about police presence? Can police actively monitor police presence and identify areas where that presence may be excessive?

In December of 2011, the Civil Rights Division of the US Department of Justice, in conjunction with the US Attorney's Office for the Western District of Washington, published the findings of a pattern or practice investigation of the Seattle Police Department (SPD) stemming from allegations of unconstitutional policing (US Department of Justice, 2011). The resulting Consent Decree led to the creation of the Performance Analytics & Research (PA&R) section in order to meet the research and analysis needs of the department in demonstrating compliance. The PA&R serves as a research and development arm of the SPD, and sponsors projects like the present effort in order to advance science while closing the distance between scientific discovery and practice.

This article reports on our attempts to develop and test a method for identifying potential over- and under-policing of neighbourhoods in Seattle, WA, through the analysis of AVL data, Computer-aided dispatch (CAD) log data, and attitudinal data drawn from the Seattle Public Safety Survey. In the next section, we outline the data, methods, and results of an initial development effort, highlighting some of the potential pitfalls one may encounter when working with these data, and some of the potential shortcomings of the method. We then discuss the utility of the data for accountability and management purposes and conclude with thoughts about future directions for research in this area.

Data and methods

In 2015, the Center for Open Policing sued the SPD for access to AVL data under the Washington State Public Records Act, RCW Chapt. 42.56, and won. In addition to 'approximately $30,000 in penalties, costs and fees', the SPD was forced to produce a redacted version of these data (Hyde and Ferguson, 2015), beginning a long effort to better understand operational vulnerabilities and practical uses for AVL data. After 'safing' the data[1] and delivering it to the plaintiffs, the present research team was engaged to explore other vulnerabilities and uses. The method described and tested in this article is typical of the culture of collaborative innovation established by PA&R.

Variables

In order to operationalize *police presence*, we relied on a literal indicator, AVL data, which consists of

[1] GPS tracks leading to or clustering around residences of officers with 'take home' vehicles and other sensitive locations (safe houses, critical infrastructure), were randomly redacted to eliminate the track to and visual cluster around these locations but so as not to leave a distinctive void.

time-stamped GPS 'pings' returned from police vehicles (every 6 s while the vehicle is in normal operation, and every second while the vehicle is in emergency operation). These data identify the location of police vehicles in time and space, and include time stamps, a unit identifier, and X–Y coordinates.[2] There are limitations to AVL data (officers may not always be with their vehicle, some areas will have bike and other specialized patrols, and some pings may be influenced by geography and the strength/quality of the GPS signal); however, for purposes of this analysis, AVL data are considered a reasonably proximal indicator of police presence. Working with AVL is a challenging 'Big Data' problem; for example, during a 2-week period in the City of Seattle the resulting AVL data file would consist of around 3 million records. Because of the complexity of working with this type of data, we restricted our initial efforts to a single precinct (the East precinct, which is one of five precincts and contains mixed-use commercial and high-density housing districts as well as single-housing residential areas) and a 2-week period during the month of August, 2013.[3] A 2-week period was chosen as a representative sample of police activity in the precinct across different officers and shifts. Additionally, most CompStat forums are conducted semi-monthly with a 14- and 28-day review period, which aligns this effort with a realistic use case. This resulted in a total of 372,804 records.

We operationalized the public demand for police presence as all 911 and non-911 telephone calls requesting police services, as well as alarm calls. Demand for police services is a complex construct (Laufs *et al.*, 2021) and calls for service data have well-known limitations (Klinger and Bridges, 1997) but we rely on them here as they are the only source of which we are aware for information about public calls to the police and other logged police activity that are available in a semi-detailed and contemporaneous fashion. The CAD log data include time stamps, fields describing the nature and priority of call, address, and X–Y coordinates. After removing records with no dispatch or primary unit identified, there were 3,186 CAD logs for analysis during the selected 2-week period. Sixty-three percent of the logs were classified as 911 calls ($n = 1{,}198$, or 38%), non-911 calls ($n = 713$, or 22%), and alarm calls ($n = 95$, or 3%).

The other 37% of CAD logs were classified as on-view activity, and relatively higher frequencies included preventative patrol ($n = 244$, or 8%), premise checks ($n = 236$, or 7%), suspicious persons ($n = 197$, or 6%), and traffic stops ($n = 128$, or 4%). We use these data to operationalize enforcement strategy.

We operationalized community preference using Seattle Public Safety Survey data drawn from an annual survey that is part of an ongoing initiative to establish tailored community policing plans in Seattle neighbourhoods, called the Micro-Community Policing Plans (MCPP).[4] The SPD MCPP is a collaboration between the SPD and Seattle University's Crime & Justice Research Center implemented in 2014 through a Community-Oriented Policing Services collaborative practitioner–academic grant. The initiative was implemented at a grass-roots level calling for precinct captains to work with community members to develop 'micro-community policing plans' for each of Seattle's 59 micro-communities (neighbourhoods). The MCPP consist of priorities and strategies developed through engagement between

[2] Different CAD/RMS systems geocode in different formats. Geocoding in use for the City of Seattle is a Projected Coordinate System, which is not limited by the error introduced by spherical projections.

[3] We recognize that these data are somewhat dated, however, as previously noted AVL data are generally regarded as sensitive and can be difficult to obtain. These data were available for the present study because they had already been produced as part of an unrelated public disclosure request. The 2-week period in August was selected because demand for police service in Seattle CAD event data tends to peak between June and September, with lows during and around the month of February.

[4] See: https://www.seattle.gov/police/information-and-data/mcpp-about.

the police and the community and through data collected through the Seattle Public Safety Survey. The Seattle Public Safety Survey instrument was developed as part of the SPD MCPP pilot, has been administered annually, and is now in its seventh year. The MCPP collaboration led by a research team comprised two faculty members and student research analysts who work in paid civilian positions assigned to one of the five Seattle Police Precincts tasked with assisting precinct captains and personnel with MCPP-related tasks and Seattle Public Safety Survey administration, data analysis, and report writing and presentations. The MCPP initiative holds annual focus groups between survey administrations with all micro-communities and recently implemented virtual community–police restorative dialogues to engage community and police in discussing the findings of the Seattle Public Safety Survey and real-time public safety concerns. The MCPP initiative and the Seattle Public Safety Survey have evolved from a grassroots implementation in 2014 to an institutionalized and integrated part of SPD practice. Seattle Public Safety Survey data is included on the public-facing data dashboard and the MCPP research team is included in SeaStat (SPD's version of CompStat) (for a detailed explanation about the survey design and methodology, see Helfgott and Parkin, 2016, 2018, 2020; Parkin and Helfgott, 2020).

The Seattle Public Safety Survey is one component of the MCPP. The Seattle Public Safety Survey is a non-probability survey translated in 11 languages administered annually since 2015. The survey is administered through broad reach-out at the precinct and micro-community levels through email, social media, media, and physical distribution of flyers citywide. The survey is intentionally designed so that all community members who live and or work in Seattle have an opportunity to take the survey. Results are statistically weighted by city demographics. Residents are asked their concerns about crime and public safety and perceptions of neighbourhood-level quality of life elements—police legitimacy, fear of crime, social cohesion, social disorganization, and informal social control. Questions about over-policing, under-policing, and police capacity are included in the survey, such as, 'On a scale from 0 to 100, with 0 being strongly disagree and 100 being strongly agree, to what extent do you agree with the following when thinking about the Seattle Police Department and its officers?' '... There is enough Seattle police officer presence in my neighborhood.' Another type of question asks, 'What, if any, are current public safety and security concerns in the neighborhood where you live and/or work?' and includes both 'over-policing of neighborhood' and 'under-policing of neighborhood' as options. For the second two questions, respondents are presented with a dichotomous option to either agree or disagree that under-policing or over-policing is a public safety concern in their neighbourhood. Data are drawn from nine micro-communities (neighbourhoods) in the East precinct—Capitol Hill, Central Area/Squire Park, Eastlake-East, First Hill, Judkins Park/North Beacon Hill, Madison Park, Madrona/Leschi, Miller Park, Montlake/Portage Bay. Survey data are available starting in 2015. Results from the 2015 survey from 7,286 respondents who live and/or work in Seattle were used in this analysis.[5] Community preference is more challenging to model since these types of data are captured in an infrequent and relatively static form, and are linked to fixed geographic aggregates, as compared to the real-time and location-specific AVL and CAD log data. We will rely on visual comparison of community preference with the other data types.

Guiding hypotheses

It stands to reason that the spatial distribution of police presence (in the form of AVL pings) should

[5] We recognize that the two-year lag between the AVL and CAD data (2013) and the survey data (2015) is not ideal, but we believe these data are still useful for conceptual/proof-of-concept purposes.

be explained by public demand (calls for service), enforcement strategy (on-view activity), and community preference (perceptions regarding police presence). If a particular location has a high level of police presence but low levels of public demand, this could be a potential indicator of over-policing. Similarly, where a concentration of police presence would be expected but not observed, an opportunity for crime control treatment may yet to be discovered. Lastly, although we cannot assess this directly due to the time-lag between the data, there should be a proximal relationship between police presence and public attitudes towards the police. Therefore, we posit three hypotheses:

> H_1: Neighbourhoods with high levels of actual police presence, low levels of public demand, and high levels of enforcement strategy will have community preferences that support less police presence.
>
> $H_{2:}$ Neighbourhoods with low levels of actual police presence, high levels of public demand, and low levels of enforcement strategy activity will have community preferences that support more police presence.
>
> H_3: Neighbourhoods with levels of actual police presence that are relatively equivalent to the levels of public demand and enforcement strategy within them will have community preferences that indicate a satisfaction with current policing levels.

Modelling strategy

Our approach was to begin by determining the location of study (in this case, the East Precinct), and limiting the data to that location and for the specific time period of study. Some initial exploration of point data and computation of spatial statistics was performed in order to understand the spatial distributions. This was followed by kernel density estimation for the three types of data being explored. We used similar parameters for the density layers in order to facilitate re-classification and potential combination. We then identify the highest density locations for the different data types, and map those locations in order to demonstrate where high presence, demand, and strategy may or may not coincide. Finally, we overlay these data layers on the community preference data. We anticipate that areas with high demand and enforcement strategy will generally have high police presence, but where there is high presence without corresponding demand or strategy there may be potential over-policing.

Results

Figure 1 depicts the AVL point data in the East Precinct of the SPD during the 2-week period. As can be seen, during the 2-week period officers drove on almost every street within the Precinct. While it might be possible to identify some degree of clustering here, it is of course very difficult to do so at the Precinct scale and there is much overlap because vehicle locations are generally constrained by roads and parking areas. The temporal component is also aggregated here. But the visualization is useful for confirming that police vehicles during the 2-week period did essentially cover the entire Precinct to some degree and the areas that were not pinged are low-density residential streets.

To visualize clustering of AVL point data, we generated a kernel density layer using 50-foot cells and a 230-foot bandwidth.[6] The density layer in Fig. 2 is symbolized using 1 standard deviation

[6] This bandwidth was determined using the default method implemented by ArcGIS, which is an adaptation of Silverman's (1986) rule and seeks a radius that is insensitive to spatial outliers but also avoids the 'ring around the points' phenomenon that can occur with too narrow a radius. The average city block length in Seattle is 240 feet, so in practical terms the default bandwidth used here is approximately one city block (which seems reasonable for understanding the density of AVL data at a particular location). One might also use local tests for spatial clustering such as the Getis-Ord Gi* statistic (see Kalinic and

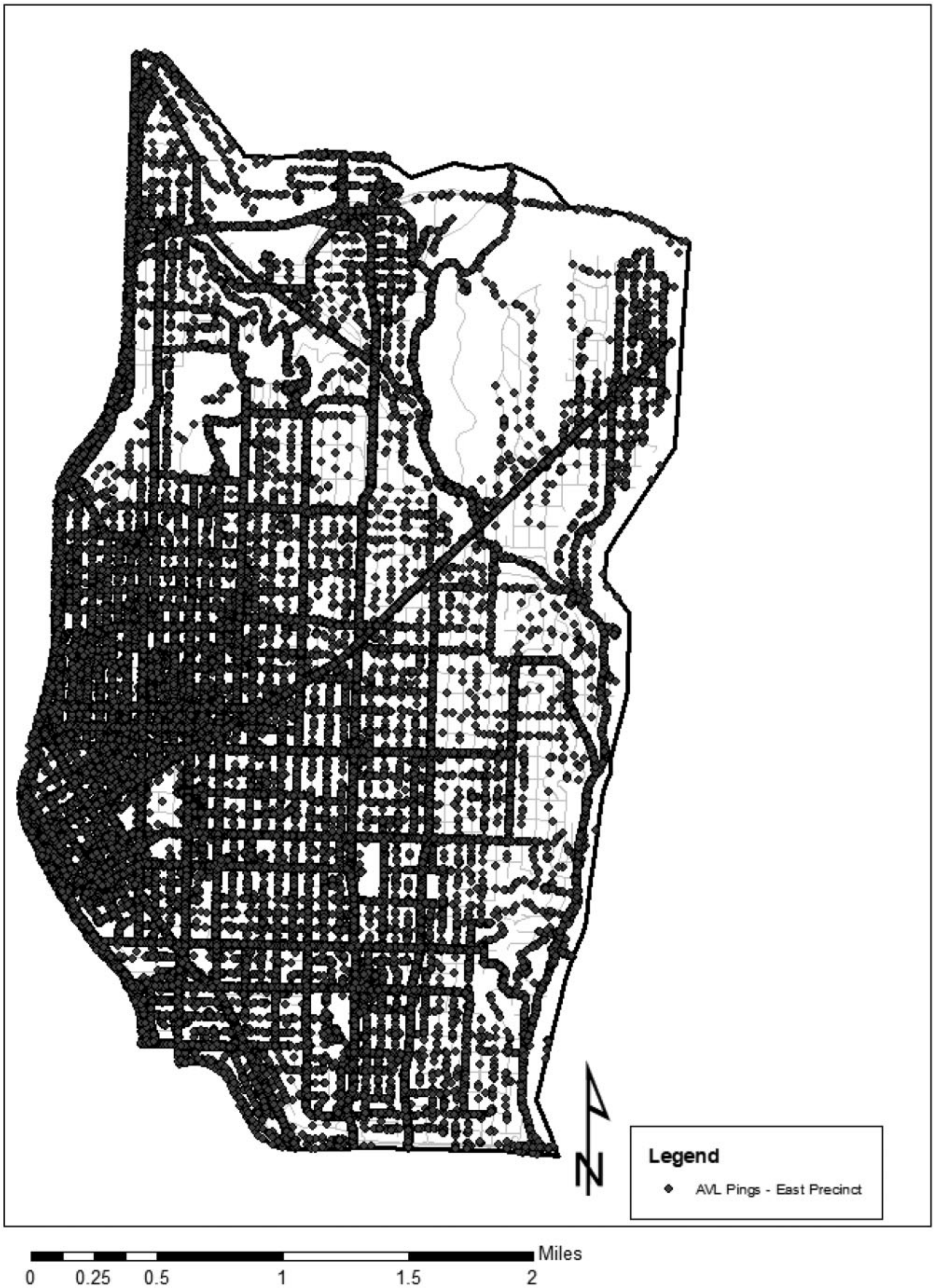

Figure 1: AVL point data, East Precinct, two week study period ($n = 372{,}804$)

(SD) breaks and depicts locations where there is greater estimated density of AVL point data. As can be seen, there are some areas where police presence is clearly more concentrated. One

Krisp, 2018), although more research is likely needed to determine appropriate applications with AVL data; we thank an anonymous reviewer for this suggestion.

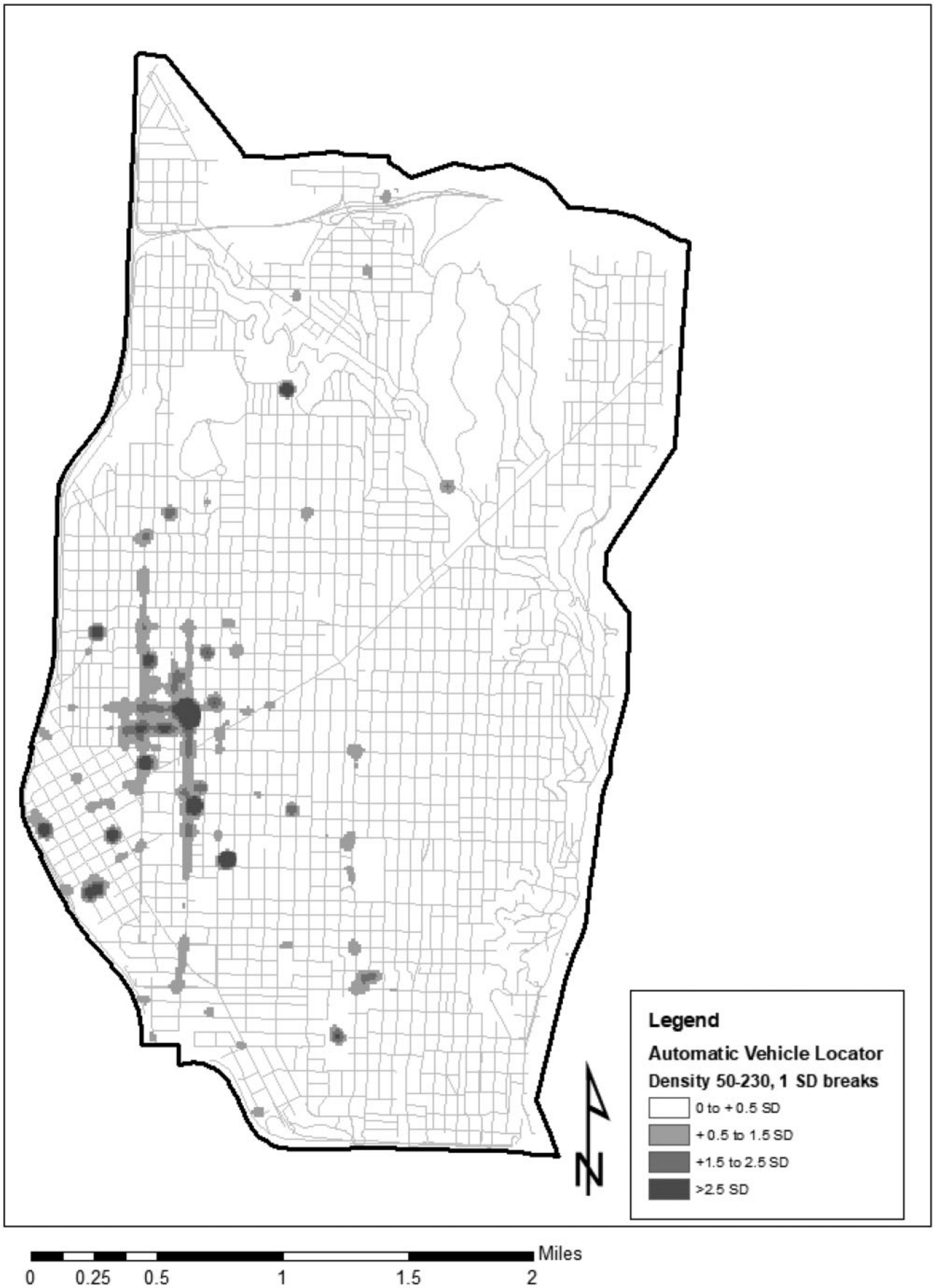

Figure 2: KDE for AVL data (50 ft cells, 230 ft bandwidth), 1 SD symbology

challenge made evident here is that police presence is 'structural' at certain locations such as the East Precinct Headquarters (which is surrounded by the largest density spot), where police vehicles are routinely departing and returning. Also, the higher density streets represent blocks with a relatively

large number of apartment complexes and retail establishments, particularly those with restaurants and establishments serving alcohol.

We next examined the calls for service and onview data using the same methods and parameters. Figure 3 presents a density layer for the calls for service data, while Fig. 4 presents a density layer for onview activity.

For each of the three layers examined thus far (police presence, public demand, and enforcement strategy), we re-classed the raster layers using 1 SD breaks and display the upper-most category (mean + 2.5 SDs) for each in Fig. 5. The colour blue is used to depict areas of high police presence; green depicts areas with high public demand; and orange depicts enforcement strategy. As can be seen, there are some areas of overlap but also several areas where the three elements are distinct. The location of the East Precinct Headquarters is noted in the map, as are three locations of high presence that are associated with hospital parking.

Red circles are positioned around four locations where there is high density of police presence, but low public demand and enforcement strategy, and no immediately obvious structural explanation (such as a police facility, hospital, or other node of police presence). Starting with the northernmost circle, further investigation reveals that this is the parking lot associated with a private elementary (K-8) school. This may indicate the presence of officers at the school for an educational or enforcement purpose; alternatively, it could be an area where officers park to eat lunch, write reports and perform other administrative tasks, or rest. Moving south to the middle-two red circled areas, the first is a parking area located behind a strip of restaurants as well as a prominent corner coffee shop, adjacent to a university campus. The second is another parking lot associated with a university recreational facility. This parking lot has historically been used by the SPD as a staging area for the management of large public demonstrations, and officers are familiar with the location as a safe place to park when writing reports or needing a break. Finally, the southernmost red circled area is the parking lot of a public middle school. Again, this may indicate the presence of officers at the school for an educational or enforcement purpose, or for alternative reasons already noted. Collectively, these areas identified as high police presence demonstrate a potential challenge with the use of AVL data, in that some masking of locations or greater selectivity may be necessary. While we have been somewhat optimistic in our assessment of police presence at these locations, it must also be acknowledged that AVL data may identify excessive or inappropriate police presence (such as sleeping while on duty or engaging in other problematic behaviours).

We now try to place these data within the context of the Seattle Public Safety Survey data that bear on community preference. Figure 6 includes the survey data (shaded areas) regarding the over-policing question, and includes the high presence, demand, and strategy layers. The percentage of residents indicated that the SPD is over-policing their neighbourhood ranges from less than 1% up to 18%, though this was only for one neighbourhood. Two neighbourhoods only had 1–3% of respondents identify over-policing as a public safety concern, and six had less than 1%. There does not appear to be much correspondence between the high activity measures and the attitudinal data. The neighbourhood with the highest percentage of residents who stated that SPD is over-policing in the neighbourhood had few indicators of increased police presence in the area. However, this result could be an artefact of the survey data as this neighbourhood had a low response rate compared to the other areas resulting in a higher sensitivity to outliers in the data. Keeping in mind the time-lag between the AVL/CAD data and the survey data, it may be the case that these measures are not time-stable and police presence may have been very different at the time the survey data were collected.

Figure 7 switches to the under-policing question (the percentage of residents indicating that the

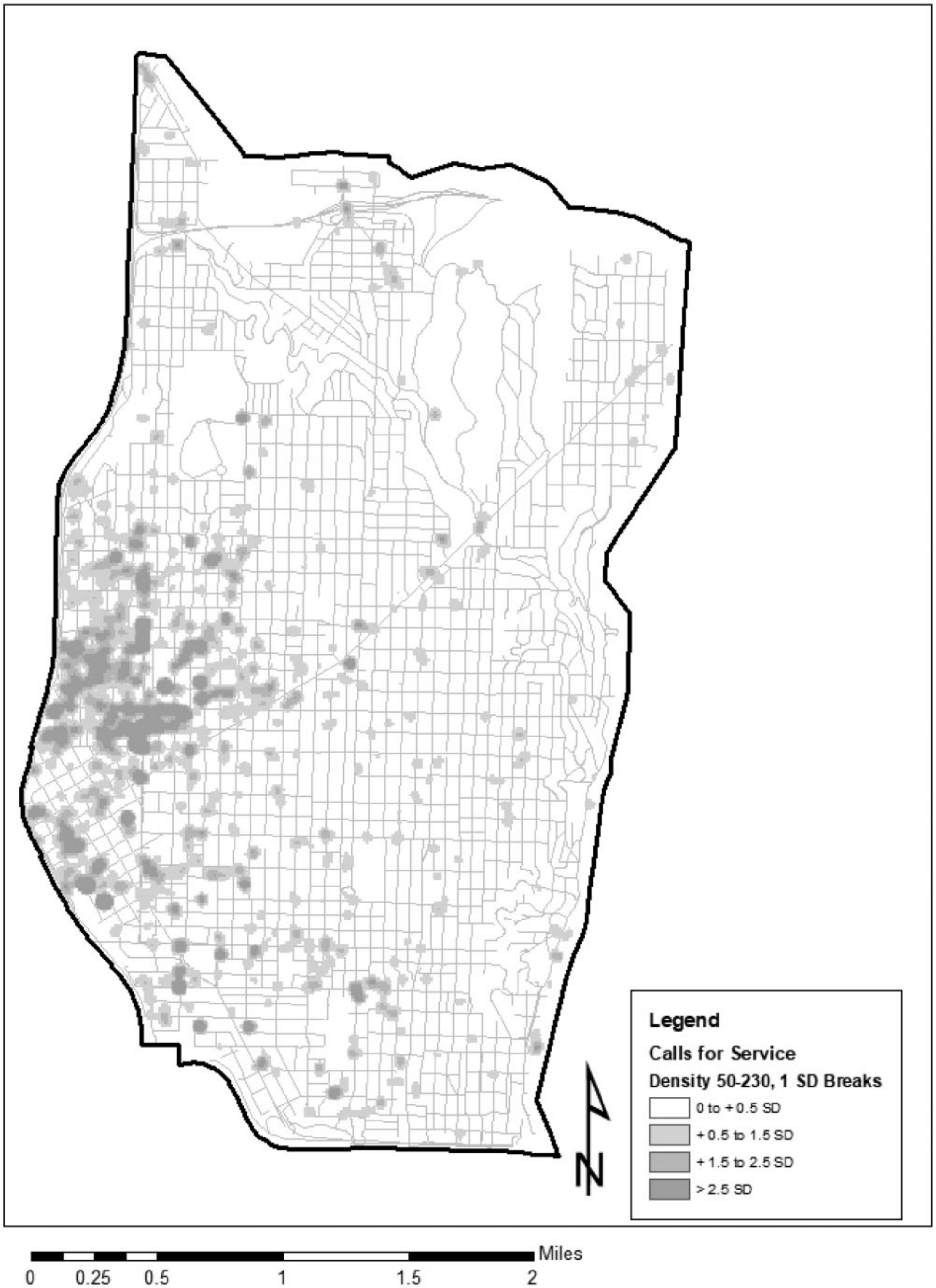

Figure 3: KDE for calls for service data (50 ft cells, 230 ft bandwidth), 1 SD symbology

SPD is under-policing their neighbourhood ranges from 10% to 45%).For these results, areas with relatively more clusters of police activity, calls for service, and onview activity have a minimum of 29% of survey respondents in those neighbourhoods indicating their communities are

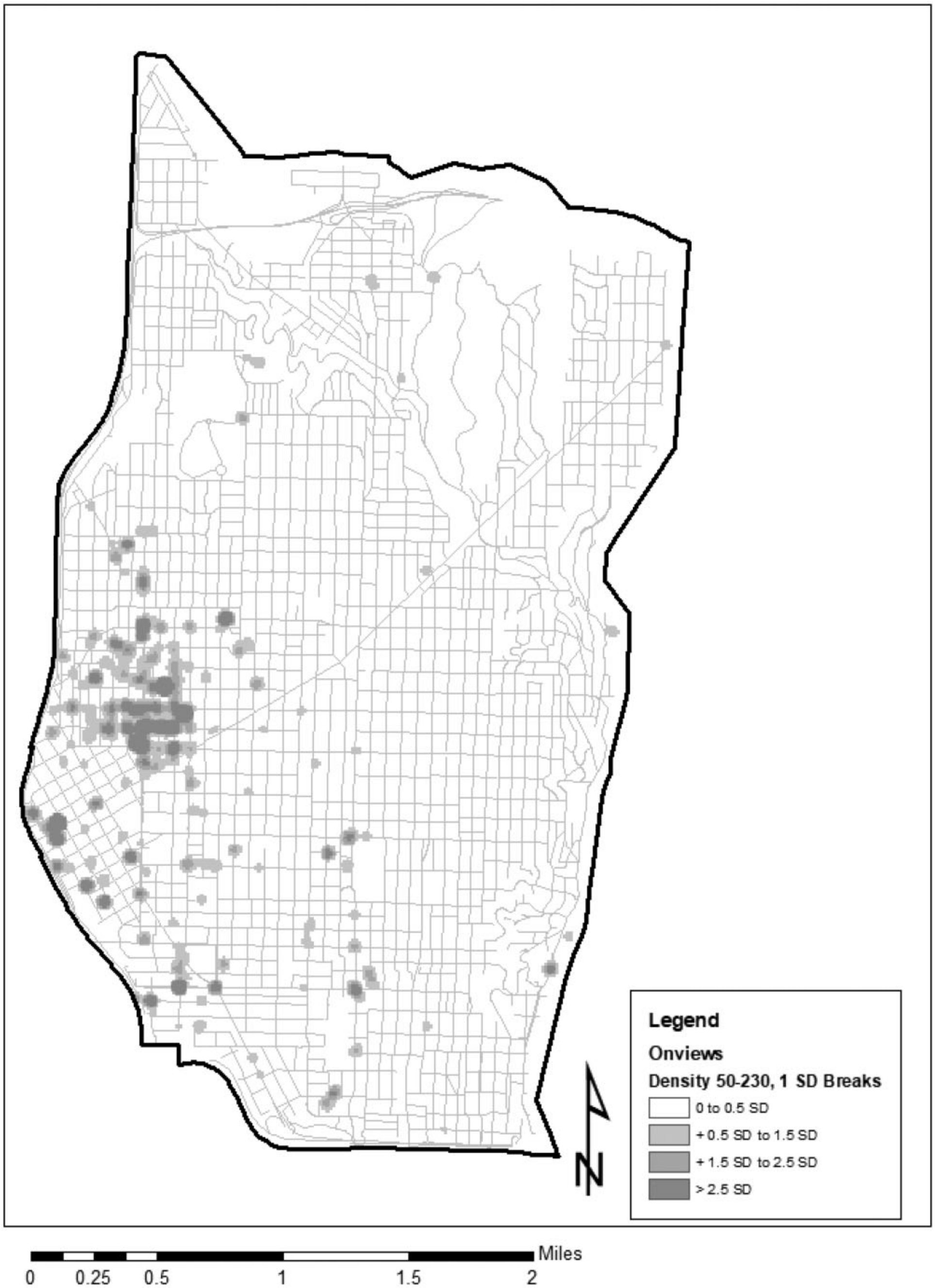

Figure 4: KDE for onview data (50 ft cells, 230 ft bandwidth), 1 SD symbology

underpoliced. Although, once again, the patterns are not uniform as communities with little or no clustering of police activity have similar survey results. Interestingly, survey respondents for communities in the northern portion of the North Precinct have little or no clustering of police activity

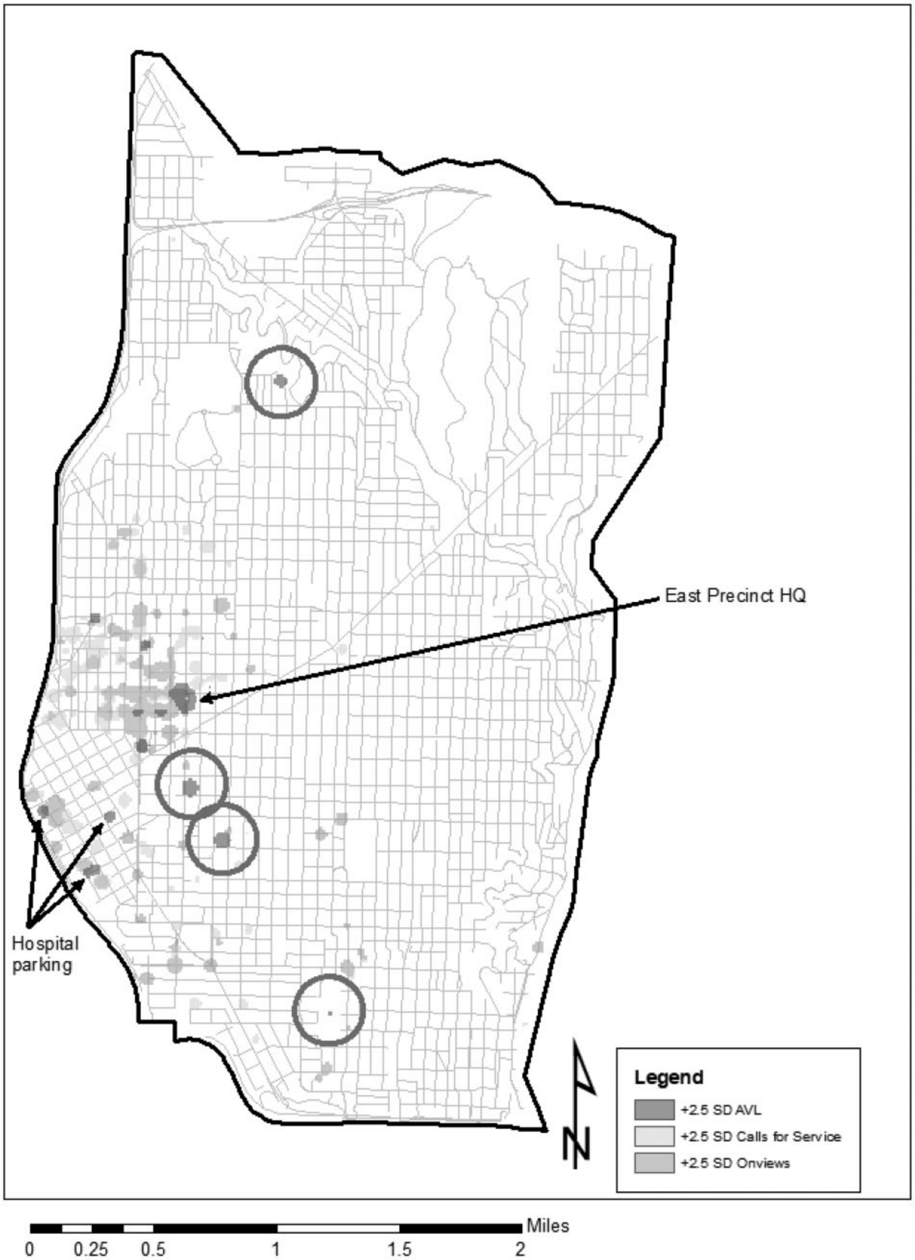

Figure 5: Overlay of re-classed KDE layers showing high presence (AVL), demand (CFS), and strategy (onviews), with red circles highlighting four areas of high presence with low demand and onview activity.

and were less likely, although not by much (22–29%), to indicate that under-policing was an issue.

Figure 8 maps the answers to the final question, which measures the average response, on a scale from 0 to 100, for how much a respondent agrees that there is enough police presence in their neighbourhood (0 no agreement, 100 full agreement). For residents indicating that there is

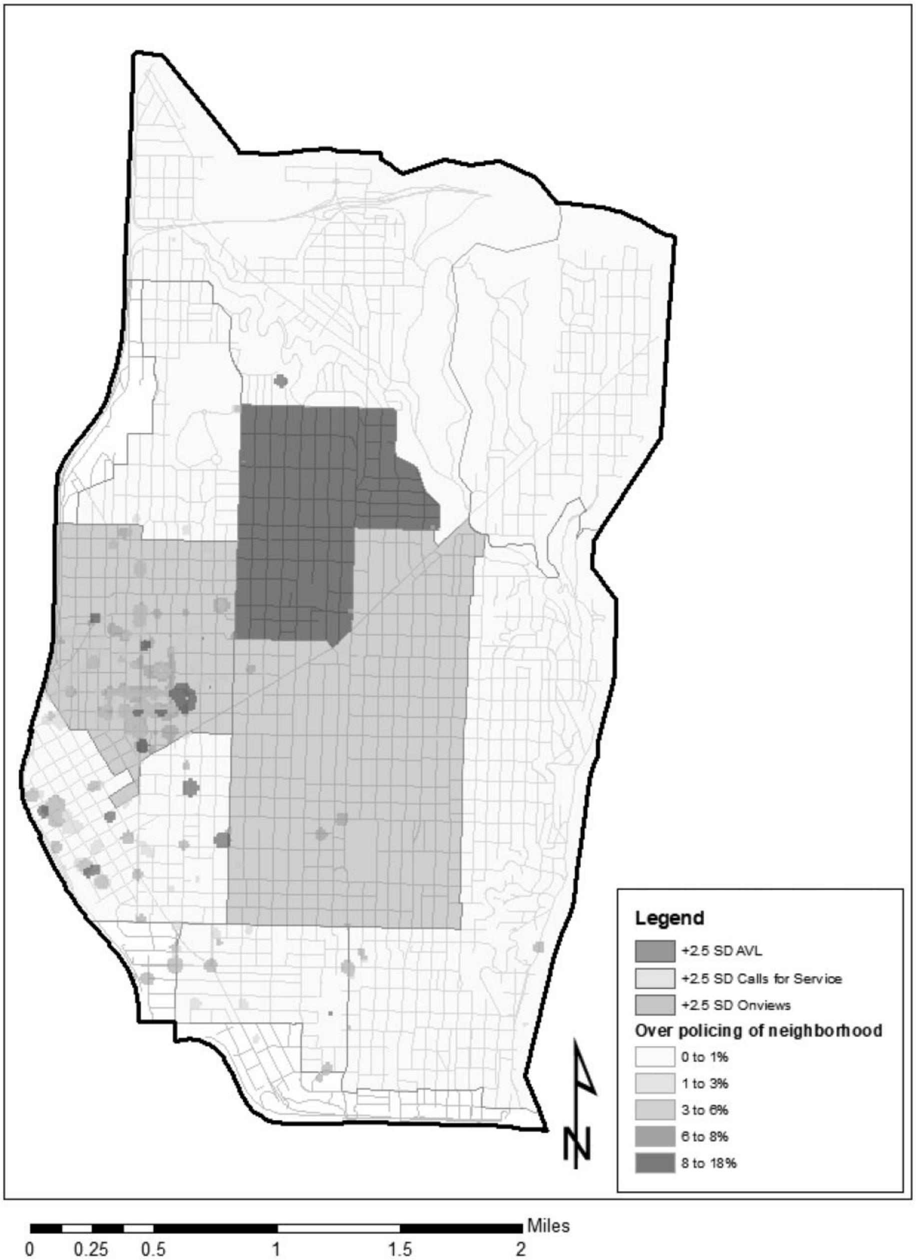

Figure 6: High presence (AVL), demand (CFS), and strategy (onviews), overlaid on level of agreement that SPD is over-policing neighbourhood.

enough SPD presence in their neighbourhood, there is a range from 35 to 60. When overlayed with the clusters of AVL, calls for service, and onview data, the neighbourhood on the west side of the East Precinct with the vast majority of clustered police activity, including calls and patrol at hospitals, shows an average survey response of 38–46 on the scale of 100 for agreeing there is

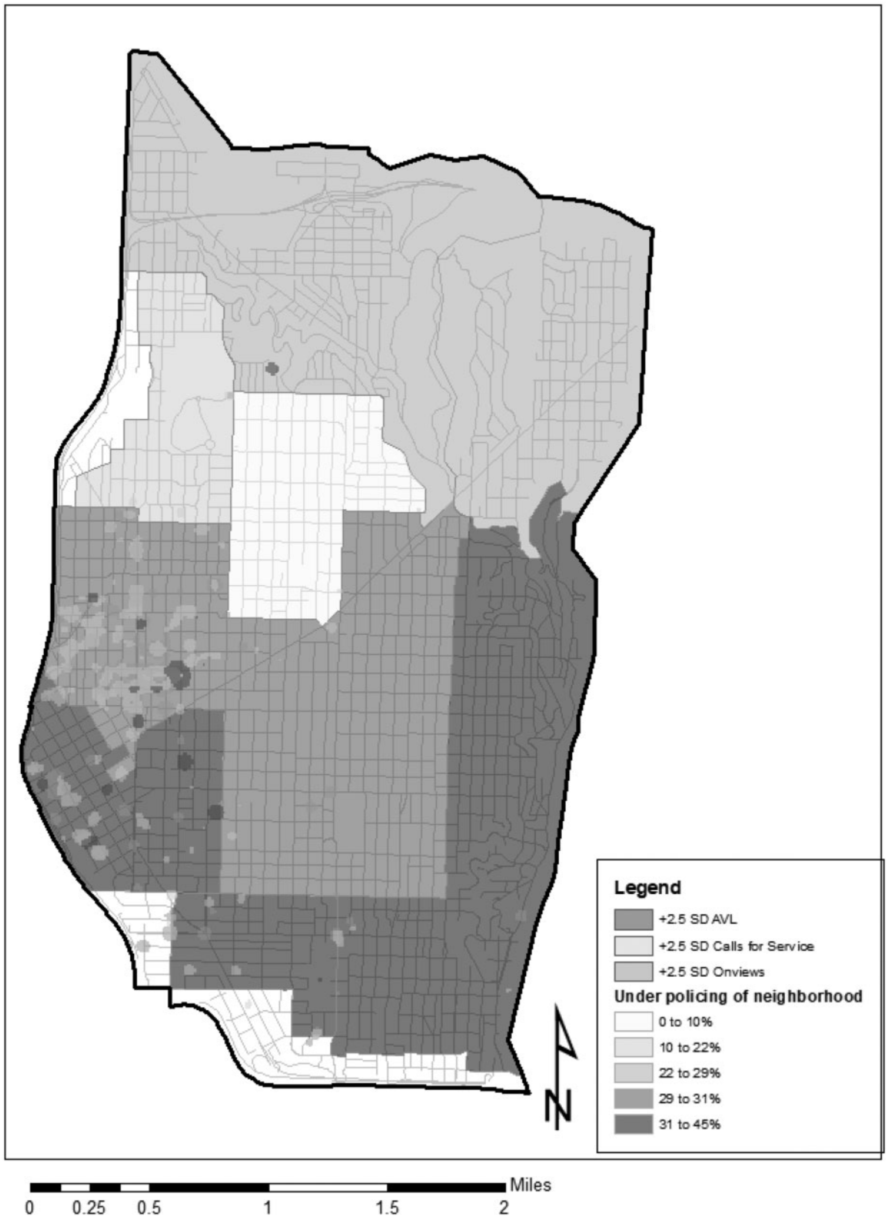

Figure 7: High presence (AVL), demand (CFS), and strategy (onviews), overlaid on level of agreement that SPD is under-policing neighbourhood.

enough police presence. The neighbourhood with the highest average of agreement with a range of 50–60 is in the centre of the Precinct with very little clustered activity, and the neighbourhood with the lowest level of agreement has clusters of calls for service and onview activity and limited AVL clusters. Also, although not in the highest tier, the neighbourhood with the highest

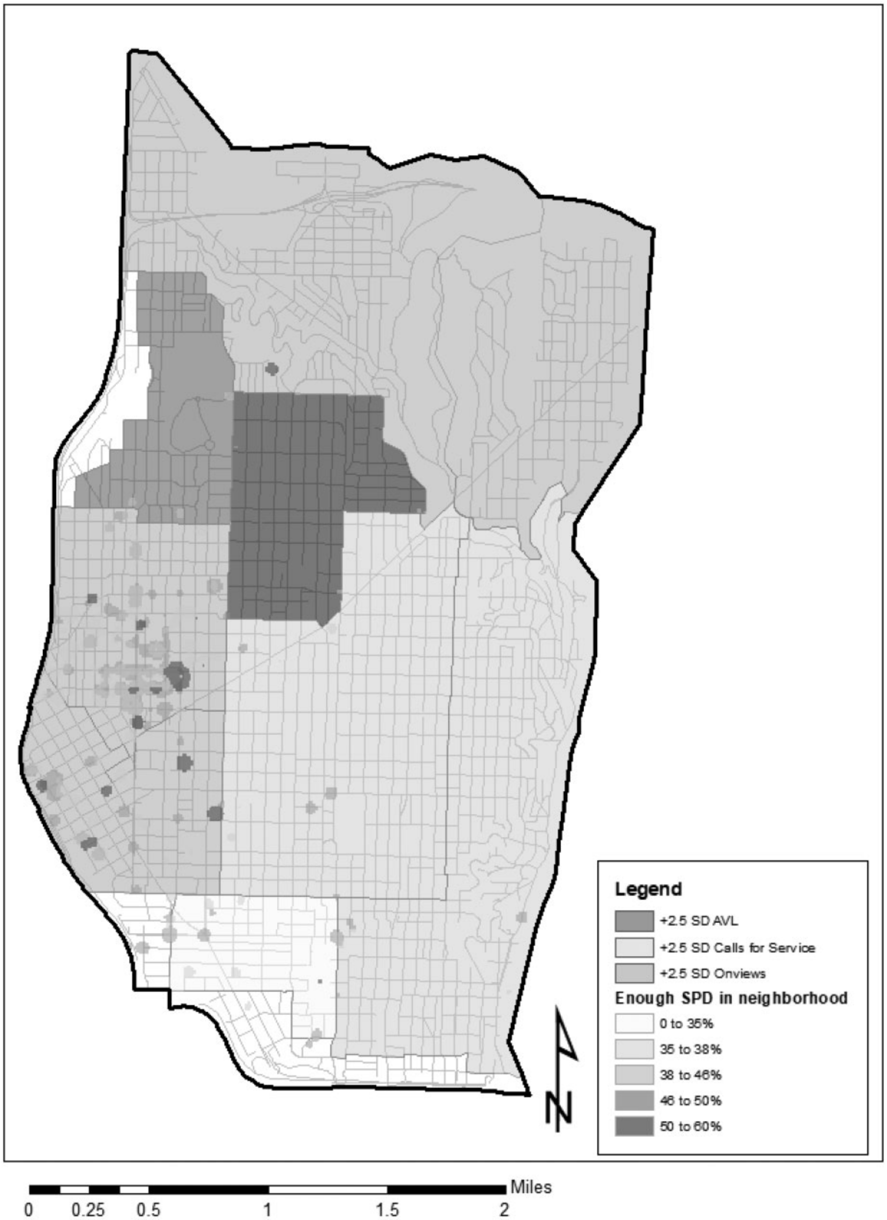

Figure 8: High presence (AVL), demand (CFS), and strategy (onviews), overlaid on level of agreement that there is enough Seattle police officer presence in neighbourhood.

density of activity for AVL, calls for service, and onview activity has average level of agreement that there is enough SPD presence in their neighbourhood.

Discussion

Returning to our hypotheses, we find weak support for our first hypothesis that stated

neighbourhoods with high levels of actual police presence, low levels of public demand, and high levels of enforcement strategy will have community preferences that support less police presence. Partly, this is because there were no communities that fit these criteria. In fact, specific to over-policing, most communities had 0–1% of respondents selected over-policing as a public safety concern. Two communities had 3–6% identify over policing as a concern. Interestingly, in one these communities, there were many clusters of calls for service, onview activity, and police presence. Perhaps elevated police presence, regardless of need, impacts the public's perception of whether the community is over-policed. In the second community, there were few clusters specific to onview activity and calls for service with no clusters of police activity. This community, however, has historically been the home to the city's largest Black population. One hypothesis to explain this could be that regardless of what police activity occurs, a history of negative relationships with law enforcement in the community drives perceptions of whether a community believes they are being over-policed, regardless of the actual activity on the ground.

We find partial support for the second hypothesis that stated neighbourhoods with low levels of actual police presence, high levels of public demand, and low levels of enforcement strategy activity will have community preferences that support more police presence. There are several neighbourhoods that have clusters of calls for service and onview activity, with no or few clusters of police activity and had 31–45% of respondents state that under policing in their community was a public safety concern. Although, there also were neighbourhoods with 31–45% of respondents stating under-policing was a public safety concern with several clusters of police activity—albeit most of these were hospital emergency rooms. The neighbourhood with the most clusters of police activity not related to hospitals or the precinct headquarters also had the most number of calls for service and 29–31% of respondents identifying under-policing as a public safety concern. These results also provide insight into the final hypothesis, which stated that neighbourhoods with levels of actual police presence that are relatively equivalent to the levels of public demand and enforcement strategy within them will have community preferences that indicate a satisfaction with current policing levels. In some ways, the types of neighbourhoods that have a disproportionate number of calls for service, onviews, and police activity, show that under-policing, not over-policing, is a concern for residents when high levels of criminal activity are occurring, regardless of the amount of police presence or how proactive they are when in the community.

Taken together, these results support the utility of a real-time data analysis tool that can map law enforcement activity, calls for service, onview activity, and public perceptions of whether their communities are being over- or under-policed—both of which if they do not match the expectations of the public can negatively impact police legitimacy and trust. The data show that although not uniform across all neighbourhoods, some communities aligned with our expectations, that perceptions of over- or under-policing could be explained by the amount of potential criminal and police activity. However, other communities did not support the hypotheses. For example, one community, which has historically been the home of the city's largest Black community had relatively higher levels of perceptions of being over-policed, although also having relatively high levels of feeling under-policed. In other cases, affluent communities with no clustering of calls for service, onviews, or police activity had relatively high levels of respondents perceiving the community was over-policed. These results could present evidence that in some cases, criminal and police activity can drive the public's perception about whether they are being under- or over-policed, but in other neighbourhoods, it is the expectations of the

community, regardless of the reality of crime and policing, that drives these perceptions.

Business intelligence in the policing context does not need to focus, strictly, on accountability. Although accountability applications tend to directly address issues of trust and legitimacy, an agency that learns to engage in real-time or near-real-time data insights would naturally promote a sense of deliberate, purposeful management, if not outright confident control of forces. CompStat, or as it is referred to in the City of Seattle, 'SeaStat,' and the management tools (e.g. reports, dashboards, intelligent decision support processes) used to monitor and respond to emergent patterns in the operation, are designed to promote engagement on the part of managers and commanders, by fostering a sense of 'data curiosity.' In this context, insights do not definitively identify problems; rather, insights generated from the distillation of large volumes of noisy data are intended to direct a user to dig deeper and understand the observation.

Within the context of processed AVL, as discussed above, the implications are myriad. For instance, an apparent overconcentration of presence might indicate over-policing, as hypothesized, or a potential operational security vulnerability (e.g. ambush risk). Diagnosing the underlying disease, from observations of the symptoms, aids in accepting the corrective action. This process of acceptance can only be developed through constant engagement. This is the philosophy underlies the CompStat model and is further enhanced by constant contact and ever more sophisticated methods of analysis.

The utility of the SPD MCPP initiative and the Seattle Public Safety Survey in examining over-policing in neighbourhoods is noteworthy. The Seattle Public Safety Survey was designed as a collaborative academic–practitioner initiative to collect annual data to inform the SPD MCPP and to increase community–police engagement and public safety. The longevity and institutional integration of the MCPP and the ongoing annual Seattle Public Safety Survey offer Seattle rich community perception data not available in other cities. Ongoing measurement of over-policing in neighbourhoods is one of the many ways the Seattle Public Safety Survey data can be used to direct police resources and city services to address quality of life elements of public safety in Seattle.

Limitations and future research

Several limitations of this research should be noted. First, the data only measure police activity over a limited time period of 2 weeks and static survey results. Also, the survey data, which is a non-probability sample taken across the city during a 45-day period, was collected 2 years after the AVL, calls for service, and onview activity data. Although we believe there is still value for conceptual/proof-of-concept purposes, this time difference could have an impact on the results as the policing and crime activity occurring within a respondent's neighbourhood immediately prior to or during the period that the survey was available could be different than the activity capture 2 years prior. Future research should attempt to collect and utilize data that is collected contemporaneously or the police activity and calls for service data immediately prior to the survey data being collected. Future research should expand to larger geographic areas wither including more of Seattle or additional jurisdictions and explore additional spatial socio-demographic and criminal justice data. There is much work to be done in understanding better ways to model AVL data and understanding spatial correlates.

In addition, our community perceptions of over- and under-policing were based on community responses to a citywide survey. Perceptions of policing could be impacted by the timing of the survey, the survey population, and the recentness of police contact with the survey respondent. Therefore, future research should develop ways to measure community perceptions of over- and

under-policing that may not be as dynamic and could provide stable estimates across time. With the important caveat about time lag in the present study, when a disconnect is observed between perception and police data this can serve as a useful mechanism to initiate discourse with the communities served and try to find out how policing can be better tailored in those communities, or to identify where greater outreach may be necessary.

We should also note that it is reasonable to raise questions about the broader practicality of the approach we propose to assessing over- and under-policing—that is this something for which we would anticipate widespread use given the resources and capabilities available in most police departments? While larger departments having crime analysts engaged in mapping would likely have the technical resources and skills or could reasonably acquire them, it is unlikely that smaller departments would have those resources, the data, or necessarily the need for this type of analysis. However, as research develops on use of AVL data for these purposes, and if it proves useful, then to the extent that the data are available it is possible that common Computer Aided Dispatch/Records Management System vendors could incorporate dashboard tools into their systems to integrate AVL and CAD data, which would make broader use more practical.

Conclusion

On balance, this study suggests that there is great potential to learn about police behaviour as well as the effects of police presence and public attitudes towards the police through analysis of AVL records and other GPS-based monitoring data. To be sure, there are reasonable privacy and safety concerns over the access and use of these types of data. But as the utility of the data become more apparent through research and development—and importantly, demonstration to police executives—we believe it will be possible to address and satisfy those privacy and safety concerns. There is great promise for addressing the problem of over- and under-policing and restoring or enhancing public trust and confidence in the police, and that alone should serve as a strong incentive for police researcher–practitioner partnerships aimed at exploring these data.

Beginning in the fall of 2021, the SPD will be one of the first police services in the United States to engage a measure of police performance other than the crime rate. The Equity Accountability and Quality forum, modelled after CompStat, will begin to use an operationalization of the method reported here in the form of hot and cold spots of policing, by precinct and watch. This programme is designed to foster a culture of continuous improvement around a more sophisticated approach to measuring the direct and indirect impacts of policing, and will attempt to bring full circle the cycle of innovation exemplified by this practitioner–academic collaboration.

References

Ben-Porat, G. and Yuval, F. (2012). 'Minorities in Democracy and Policing Policy: From Alienation to Cooperation.' *Policing & Society* **22**(2): 235–252.

Blanes i Vidal, J. and Mastrobuoni, G. (2018). *Police Patrols and Crime*. Discussion Paper No. 11393. Bonn, Germany: IZA Institute of Labor Economics.

Boehme, H. M., Cann, D. and Isom, D. A. (2020). 'Citizens' Perceptions of over- and under-Policing: A Look at Race, Ethnicity, and Community Characteristics.' *Crime & Delinquency* **68**: 1–32.

Chalfin, A., Hansen, B., Weisburst, E. and Williams, M. (2020). *Police Force Size and Civilian Race*. NBER Working Paper No. 28202. Cambridge, MA: National Bureau of Economic Research.

DeAngelo, G., Toger, M. and Weisburd, S. (2020). *Police Response Times and Injury Outcomes*. Discussion Paper No. 14536. London: Centre for Economic Policy Research.

Goldsmith, A. and Harris, V. (2012). 'Trust, Trustworthiness and Trust-Building in International Policing Missions.' *Australian & New Zealand Journal of Criminology* **45**(2): 231–254.

Gibson, C., Slothower, M. and Sherman, L. (2017). 'Sweet Spots for Hot Spots? A Cost-Effectiveness Comparison

of Two Patrol Strategies.' *Cambridge Journal of Evidence Based-Policing* **1**(4): 225–243.

Helfgott, J. and Parkin, W. (2016). *Seattle Police Department's Citywide 2015 Seattle Public Safety Survey Results*. Unpublished report. http://www.seattle.gov/Documents/Departments/Police/Reports/2015%20Seattle%20Public%20Safety%20Survey%20Results_CITYWIDE.pdf (accessed 22 October 2021).

Helfgott, J. B. and Parkin, W. (2018). Seattle police department's micro-community policing plans evaluation. US Department of Justice. *Office of Community Oriented Policing Services Cooperative Agreement Number 2014-CKWX-K044* www.seattle.gov/Documents/Departments/Police/Reports/2018_Seattle_PD_MicroCommunities.pdf (accessed 22 October 2021).

Helfgott, J. B. and Parkin, W. S. (2020). 'Seattle Public Safety Survey Probes Safety for Everyone, Including People Who Are Unhoused.' *Real Change* **27**(46): 2–5.

Hough, M. (2012). 'Researching Trust in the Police and Trust in Justice: A UK Perspective.' *Policing and Society* **22**(3): 332–345.

Hyde, D. and Ferguson, A. (2015). Why Seattle programmers sued the police for GPS data. *KUOW.org*. https://www.kuow.org/stories/why-seattle-programmers-sued-police-gps-data (accessed 22 October 2021).

Kalinic, M. and Krisp, J. (2018). *Kernel density estimation (KDE) vs. hot spot analysis - detecting criminal hot spots in the City of San Francisco*. Paper presented at Agile 2018 - 21st Conference on Geo-information Science, Lund, Sweden.

Klinger, D. A. and Bridges, G. S. (1997). 'Measurement Error in Calls-For-Service as an Indicator of Crime.' Criminology **35**(4): 705–726.

Klinger, D. (1997). 'Negotiating Order in Patrol Work: An Ecological Theory of Police Response to Deviance.' *Criminology* **35**(2): 277–306.

Kochel, T., Burruss, G. and Weisburd, D. (2015). *St. Louis County Hot Spots in Residential Areas (SCHIRA) Final Report: Assessing the Effects of Hot Spots Policing Strategies on Police Legitimacy, Crime, and Collective Efficacy*. Washington, DC: US Department of Justice.

Koper, C. (1995). 'Just Enough Police Presence: Reducing Crime and Disorderly Behavior by Optimizing Patrol Time in Crime Hotspots.' *Justice Quarterly* **12**(4): 649–672.

Laufs, J., Bowers, K., Birks, D. and Johnson, S. (2021). 'Understanding the Concept of 'Demand' in Policing: A Scoping Review and Resulting Implications for Demand Management.' *Policing & Society* **31**(8): 895–918.

Lum, K. (2017). 'Limitations of Mitigating Judicial Bias with Machine Learning.' *Nature Human Behaviour* **1**(7): 1–1.

Mitchell, R. (2017). 'Frequency versus Duration of Police Patrol Visits for Reducing Crime in Hot Spots: Non-Experimental Findings from the Sacramento Hot Spots Experiment.' *Cambridge Journal of Evidence Based-Policing* **1**(1): 22–37.

Parkin, W. S. and Helfgott, J. B. (2020). All eyes on public safety: A look at Seattleites' concerns about their neighborhoods. *The Seattle Times*. https://www.seattletimes.com/opinion/all-eyes-on-public-safety-a-look-at-seattleites-concerns-about-their-neighborhoods/ (accessed 22 October 2021).

O'Brien, G. (2021). 'Racial Profiling, Surveillance and over-Policing: The over-Incarceration of Young First Nations Males in Australia.' *Social Sciences* **10**(2): 68–10.

Perry, B. (2006). 'Nobody Trusts Them! Under-and Over-Policing Native American Communities.' *Critical Criminology* **14**(4): 411–444.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*. New York, NY: Chapman and Hall.

Telep, C., Mitchell, R. and Weisburd, D. (2014). 'How Much Time Should the Police Spend at Crime Hot Spots? Answers from a Police Agency Directed Randomized Field Trial in Sacramento, California.' *Justice Quarterly* **31**(5): 905–933.

US Department of Justice. (2011). *Investigation of the Seattle Police Department*. Washington, DC: US Department of Justice.

Wain, N. and Ariel, B. (2014). 'Tracking of Police Patrol.' *Policing* **8**(3): 274–283.

Weisburd, D., Groff, E., Jones, G. et al. (2015). 'The Dallas Patrol Management Experiment: Can AVL Technologies Be Used to Harness Unallocated Patrol Time for Crime Prevention.' *Journal of Experimental Criminology* **11**(3): 367–391.

Weisburd, S. (2021). 'Police Presence, Rapid Response Rates, and Crime Prevention.' *The Review of Economics and Statistics* **103**(2): 280–293.

Weisburd, S. (2016). *Does Police Presence Reduce Car Accidents*? Discussion Paper No. 12–16, The Pinhas Sapir Center for Development. Tel Aviv-Yafo, Israel: Tel Aviv University.

Weisburd, S. (2013). *The Effect of Police Patrol on Car Accidents*. Jerusalem: The Hebrew University.

Yapo, A. and Weiss, J. (2018). Ethical implications of bias in machine learning. *Paper presented at the Hawaii International Conference on System Sciences*. DOI: 10.24251/HICSS.2018.668